UNIVERSITEIT GENT

FACULTY OF SCIENCES

Do patients die from or with infection?

Finding the answer through causal analysis of longitudinal intensive care unit data.

Maarten Bekaert

Universiteit Gent
Faculteit Wetenschappen
Vakgroep Toegepaste Wiskunde en
Informatica

Promotoren: Prof. Dr. Stijn Vansteelandt[1]
Prof. Dr. Dominique Benoit[2]

[1] Universiteit Gent
Faculteit Wetenschappen
Vakgroep Toegepaste Wiskunde en Informatica
Krijgslaan 281, S9, B-9000 Gent, België

[2] Universitair ziekenhuis (UZ) Gent
Vakgroep Intensieve zorgen
De Pintelaan 185, B-9000 Gent, België

Dit werk kwam tot stand in het kader van een specialisatiebeurs van het IWT-Vlaanderen (Instituut voor de aanmoediging van Innovatie door Wetenschap en Technologie in Vlaanderen).

Proefschrift tot het bekomen van de graad van
Doctor in Statistische Data Analyse
Academiejaar 2011-2012

Table of Contents

1 General introduction

1.1 Introduction

Nosocomial infections form a major public health problem in the Western world and accurate knowledge of the *causal impact* of these infections on mortality is important, not only to be able to assess the severeness of these infections but also to understand the importance of early treatment and to be able to estimate the high costs which they put to the community as a result of an extended length of stay and extra cost by hospitalization. Despite decades of research in the medical literature, assessment of the attributable mortality due to nosocomial infections in the intensive care unit (ICU) remains controversial, with several studies describing effect estimates ranging from being neutral to extremely risk increasing (Carlet, 2001; Melsen et al., 2009; Muscedere, 2009). This large variability has been primarily explained by differences in the patient population under study (case-mix), the definition and diagnosis of several nosocomial infections as well as the use of various definitions of attributable mortality contributes to that variation (Timsit et al., 2011). In this thesis we focus on methodological shortcomings at the level of the data analysis. A major step forward in the analysis of ICU-data with the application to nosocomial infections has been made by Wolkewitz et al. (2008, 2009), Nguile-Makao et al.

(2010) and Beyersmann et al. (2008) in which they introduced nosocomial infections as a time dependent covariate to reduce time dependent bias (Van Walraven et al., 2004). But, as investigators have to rely on observational data alone, careful adjustment for confounding by severity of illness is required to disentangle the complex relationship between infection and mortality. This is because infection is essentially a complication of underlying critical illness so that patients who acquire infection tend to be more severely ill than patients who do not.

In order to accommodate the challenges related to the estimation of the attributable mortality of nosocomial infections in the ICU, a new method from the field of *causal inference* was developed in **chapter 2** and applied in the ICU-literature to find an answer to the question if patients die *from* or *with* infection in **chapter 3**. In this process, close collaboration with physicians who were convinced of the need for more sophisticated statistical methods, was very important. Instead of further improving the previous methodology we additionally explored two other topics within the field of *causal inference*. This change of focus was motivated by the fact that the weighted estimator for the causal effect of infection, involves a model selection process in which it is unclear how to decide which covariates to select. This model selection process can be important, considering that weighted estimators can be highly inefficient due to highly variable weights. Therefore, in **chapter 4** we moved our focus to *model selection and model misspecification in causal inference* where we describe a model selection procedure in terms of the quality of the effect estimate of interest. In **chapter 5** we developed insights on to whether inverse probability weighting can be avoided altogether, using the simpler problem of simultaneous estimating the *natural direct and indirect effects* as a case study.

This thesis is a combination of stand alone papers with additional extensions and further insights obtained during my four years of research. *Chapter 2* was published in *Lifetime data analysis* (Bekaert et al., 2010b). The developed methodology is applied in **chapter 3** and published in *The American Journal of Respiratory and Critical Care Medicine* (Bekaert et al., 2011), the world's

leading intensive care unit journal. **Chapter 4** appeared in *Statistical Methods in Medical Research* (Vansteelandt et al., 2010a) where my contribution was mainly on the model selection part of the paper. Finally the work on the estimation of natural direct and indirect effects resulted in two additional papers which are currently under revision. The general idea which was independently developed by Theis Lange is submitted as a *proof of concept paper*. In a follow up paper we propose additional estimators for the natural direct and indirect effect which can be simply implemented in standard software.

In the remainder of this chapter we give a general introduction to some of the primary topics related tho my thesis. The aim is to give a concise explanation and to provide enough information to enable the reader to understand the other chapters.

1.2 Causal inference

In this section I give a brief summary of some basic ideas from *causal inference*. A more detailed description of these ideas can be found in some very interesting papers which I strongly recommend to those who whish to learn more about the topic. A nice paper on the definition of causal effects is written by Hernán (2004). The problem of time varying confounding and the use of *marginal structural models* MSMs is wel described in Robins et al. (2000) with an application in Hernán et al. (2000). For an introduction to causal DAGs see Greenland et al. (1999a)

1.2.1 Defining causal effects

Let us think about a patient who has been admitted to the ICU, acquires an infection and dies within the ICU. For simplicity we will ignore the timing of the events but discuss this topic in more detail in **chapter 2**. Imagine that we can somehow know, that the patient would still be alive had he not acquired infection. In that case, infection caused the patient's death and infec-

tion had a *causal effect* on the patients'survival. This example illustrates how we reason about causal effects: *in terms of a comparison of two outcomes for the same patient where an action is taken and withheld, every thing else being the same.* When the two outcomes differ, we say that the action has a causal effect on the outcome. Throughout this thesis I will refer to the action as the exposure or treatment.

Consider now a dichotomous exposure variable A (1: infected, 0: not infected) and a dichotomous outcome variable Y (1: death, 0: survival). Now, we use $Y(a)$ to denote the outcome variable that would have been observed under the infection state a. For our example patient, infection A has a *causal effect* on the outcome Y because $Y(1) \neq Y(0)$. In the causal inference literature, the variables $Y(1)$ and $Y(0)$ are referred to as *potential* or *counterfactual outcomes* because they represent situations that can be potentially observed, but may not actually have occured (counter to the fact).

To link these counterfactual outcomes to the observed data we will assume that for the infected ($A = 1$) patient in the example the counterfactual outcome $Y(1) = 1$ is equal to his observed outcome $Y = 1$. This brings us to the first key assumption in causal inference, which we will make throughout: the *consistency assumptions* (VanderWeele, 2009) which formally states that if $A_i = a$, then $Y_i(a) = Y_i(A) = Y_i$ or a subject with observed treatment A equal to a, has observed outcome Y equal to his counterfactual outcome $Y(a)$. In other words, it assumes there exists an intervention whereby infection is acquired (prevented) for all patients and that this intervention is non-invasive, in the sense that for patients who would naturally acquire (prevent) infection, the same event time and status would be observed under that intervention as was naturally observed.

Because, for every patient admitted to the ICU we thus only observe one outcome, *individual causal effects* cannot be identified and we therefore focus on the *average causal effect* in

a population of individuals. Take e.g. a random subset of 100 patients admitted to the ICU of the Ghent University Hospital in 2011 as our population of interest. Imagine that we know both counterfactual outcomes for all of the patients. So for each patient we have $Y(1)$ which is the outcome that would have been observed if he were infected and $Y(0)$ if he remained infection free. Infection has an *average causal effect* on mortality if $E[Y(1)] \neq E[Y(0)]$ in our population of ICU patients. In the real world, we only observe one counterfactual outcome because a patient is either infected or not. For every patient the observed outcome is equal to Y and the observed infection status is A. We say that infection A is *associated* with mortality Y if $E[Y|A = 1] \neq E[Y|A = 0]$ or if there is a difference in mortality in the infected $A = 1$ and the infection free $A = 0$ patients which form *two disjoint subsets of the population*. In contrast to the *causal effect* of infection which is defined as a difference in mortality in *the entire population under two different hypothetical infection states* a (1 = infected or 0 = not infected).

Let us put aside the practical and ethical concerns and think about a study (experiment) in which we *randomly* (by flipping a coin) divide a group of patients in the ICU into two groups (group 1 and 2). Imagine that its possible to infect all patients of group 1 and prevent infection for all patients in group 2. Afterwards we look at the mortality differences between both groups. Because group membership is randomized it is irrelevant for the value of $E[Y|A = a]$ which particular group was infected. We say that the patients in both groups are *exchangeable* because the mortality in group 1 would have been the same as the mortality in group 2 had the patients in group 2 been infected instead of those in group 1. More formally this means that $Y(a) \coprod A$ for both $a = 1$ and $a = 0$, which denotes that $Y(a)$ is independent from A. In that case the association $E[Y|A = 1] - E[Y|A = 0]$ equals the causal effect $E[Y(1)] - E[Y(0)]$. So in perfectly randomized experiments, *association* is equal to *causation*.

In practice it is not possible to conduct such experiment and we need to rely on *observational* data. The question now is how such observational data can be used for causal inference. We

know that acquiring infection in the ICU will not be a random process but depends on the patient's underlying severity of illness L. In order to infer the causal effect of infection on mortality we need to assume that $Y(a) \coprod A|L$ which is usually referred to as the *no unmeasured confounders assumption* as it effectively states that variables reflecting the patient's underlying severity of illness L are the only confounders of the association between infection and mortality. If this assumption holds then the conditional association between infection A and mortality Y, given L, will reflect the causal effect of infection within levels of L. Because of the *consistency assumption* $E[Y|A = 1, L] - E[Y|A = 0, L]$ equals $E[Y(1)|A = 1, L] - E[Y(0)|A = 0, L]$. Further by the *no unmeasured confounders assumption* this equals $E[Y(1)|L] - E[Y(0)|L]$ or $E[Y(1) - Y(0)|L]$ which reflects the causal effect of infection within levels of L. This implies that inferring causal effects from the observational data will require some covariate adjustment by L. In standard regression analysis this is done by including the covariates L in the model. More sophisticated methods use weights to adjust for confounding.

1.2.2 Direct and indirect causal effects

An important problem within both epidemiology and many social sciences is to decompose the effect of given treatment or exposure into different causal pathways and quantify the importance of each of these pathways. In view of this, it is very important to distinguish between common causes of the exposure and the outcome (confounders), and so called *mediators*, which are variables that are possibly affected by the exposure and lie on the causal path (in the direction of the arrows) from the exposure to the outcome. In this section we introduce the concept of direct and indirect causal effect using a counterfactual approach

Let for each subject $Y(a, m)$ be the outcome that we would, possibly contrary to the fact, have observed for that subject had the exposure A been set to the value a and, the mediator M was set to m. Similarly, the counterfactual variable $M(a)$ denotes the value of the mediator if, possibly contrary to the fact, the exposure A was set to a. Let C be a set of baseline confounders for the

exposure-outcome, exposure-mediator and the mediator-outcome relationship. In general, there are two types of direct and indirect effects. The *controlled direct effect* can be seen as a result from a non-invasive intervention changing the exposure A (from a to a^*) and fixing the mediator M at m and is formally defined as $E[Y(a,m) - Y(a^*,m)]$ which is the comparison of the counterfactual responses $Y(a,m)$ corresponding to different levels a, but the same level m, for all individuals. Note that the magnitude of the controlled direct effect may differ across each possible value of M. Under the assumption of no unmeasured confouders for the exposure-outcome and exposure-mediator relationship and consistency assumption this effect

$$\begin{aligned} E[Y(a,m) - Y(a^*,m)] &= E[Y(a,m)|A=a, M=m] - E[Y(a^*,m)|A=a^*, M=m] \\ &= E[Y|A=a, M=m] - E[Y|A=a^*, M=m] \end{aligned}$$

can be estimated from observed data.

The second type of effect is called the *natural direct effect*. It encodes the effect of A on Y when the mediator M takes on the value it would *naturally* have under a value a for A. Following the tradition in the causal inference literature (Hafeman and Schwartz, 2009), we will describe natural direct and indirect effects in terms of so-called nested counterfactuals, $Y(a, M(a^*))$, denoting the outcome that would have been observed if A were set to a and M to the value it would have taken if A were set to a^*. In particular, we will compare $Y(a, M(a^*))$ with $Y(a^*, M(a^*))$ to obtain a measure of the natural direct effect of changing the exposure from a to a^*. Such comparison can for instance be made in terms of an average difference within levels of covariates, $E[Y(a, M(a^*)) - Y(a^*, M(a^*))|C]$, or marginally, $E[Y(a, M(a^*)) - Y(a^*, M(a^*))]$; as a risk ratio, $P[Y(a, M(a^*)) = 1]/P[Y(a^*, M(a^*)) = 1]$, etc. Likewise we will compare $Y(a^*, M(a))$ with $Y(a^*, M(a^*))$ to obtain a measure of the *natural indirect effect*.

The expression for the *controlled direct effect* assumes (1) that $Y(a,m) \coprod A|C$ or that there are no-unmeasured confounders for the $A - Y$ relationship and (2) that $Y(a,m) \coprod M|A, C$ or that

one knows all confounders of the association between M and Y. The set C must thus contain all of the confounders of both the $A - Y$ relationship and the $M - Y$ relationship. In addition to the previous assumptions, the expressions for *natural direct and indirect effects* require that $M(a) \coprod A|C$ or that one has measured all confounders of the association between A and M. Additionally there should not exist variables L that are effects of exposure and that confound the $M - Y$ relationship so that $Y(a, m) \coprod M(a^*)|C$. In **chapter 5** we show how, under previous assumption, natural direct and indirect effects can be estimated from observational data.

1.2.3 Causal DAG's

In the previous section we used *counterfactual theory* to formalize causal effects together with the identifying conditions necessary to estimate those effects. In this section we will introduce a graphical representation of the data generating mechanism behind the observational data. As in the previous section, let A, Y and L represent infection, mortality and disease severity, respectively. Then the data generating mechanism can be represented by a *causal directed acyclic graph* (DAG) as in Figure 1.1 which consists of 4 nodes representing the random variables (A, Y, L and U) and five edges (arrows). The diagram is *directed* because the edges imply a direction in which A may cause Y but not the other way around. Further, it is called *acyclic* because there are no cycles because a variable cannot cause itself (even through another variable). In general a causal DAG is a DAG in which the arrows can be interpreted as direct causal effects, and all common causes of any pair of variable are included in the graph.

In the DAG represented in Figure 1.1, the arrow between L and A expresses that the probability of acquiring an infection may depend on the patient's underlying disease severity which is possibly affected by unmeasured variables U. The direction of the effect (harmful or protective) can not be distinguished from the DAG.

Thinking about the data generating mechanism is an important first step in the analysis of ob-

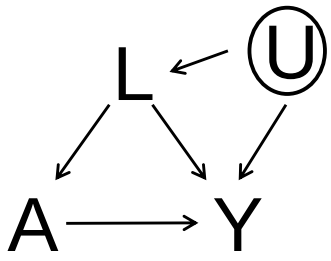

Figure 1.1: Causal DAG to illustrate the phenomena of confounding

servational data. A DAG is therefore used together with *d-seperation*: a graphical rule to verify statistical independencies. In our example, mortality status Y would be statistically independent of infection A if $f[Y|A=1] = f[Y|A=1] = f[Y]$. If we look at different paths (arrow-based route between two variables, disregarding the direction of the arrows) in the DAG a path can be either *open* or *blocked*. In the DAG represented in Figure 1.1 the path $A \rightarrow Y \leftarrow L$ is blocked because two arrowheads on the path collide in Y. Y is therefore referred as a *collider*. On the other hand, the paths $L \rightarrow A \rightarrow Y$, $A \leftarrow L \rightarrow Y$ and $A \leftarrow L \leftarrow U \rightarrow Y$ are open paths because no two arrowheads on the path collide. An open path can be blocked after *conditioning on a non-collider*. Conditioning on a collider will have the effect of opening the blocked path. Note that the path $A \leftarrow L \leftarrow U \rightarrow Y$ cannot be closed by conditioning on U because U is an unmeasured confounder.

DAG's are very useful tools in this step because they can reveal possible sources of bias. One well known sort of bias is *confounding* in which the association measure differs from the causal effects measure. A detailed explanation about confounding will be given in **chapter 4**. In general, confounding exists when the exposure and the outcome share a common cause. In our example the underlying disease severity L is a confounder for the relationship between infection A and mortality Y. More formally this means that $Y(a) \coprod A$ will not hold. When we don't adjust the analysis for L, the association measure will be a mix of the causal effect along the path $A \leftarrow Y$ and the spurious association induced by the open path $A \leftarrow L \rightarrow Y$ because

here, L is a non-collider. In order to obtain an unbiased estimate of the causal effect of infection on mortality we will need to condition or adjust the analysis for the patient's severity of illness because $Y(a) \coprod A|L$. Conditioning on L will block the path $A \leftarrow L \rightarrow Y$ and we will obtain a direct infection effect along the path $A \rightarrow Y$ by comparing infected and non-infected patients with the same underlying severity of illness L.

In the DAG in Figure 1.2, M is called a *mediator* because it lies on the causal path from A to Y.

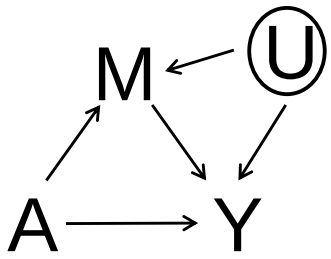

Figure 1.2: Causal DAG to illustrate the phenomena of mediation

An unadjusted analysis of the data generated under the DAG in Figure 1.2 estimates the *total causal effect* which is a combination of the *direct causal effect* along the path $A \rightarrow Y$ and the *indirect causal effect* along the path $A \rightarrow M \rightarrow Y$. Using d-separation it is clear that conditioning on M will not result in the estimation of the direct causal effect because M is a collider on the path $A \rightarrow M \leftarrow U \rightarrow Y$ and induces a spurious association.

Until now we didn't think about exposures which can vary over time. In the next section we will introduce the problem of time varying confounding and show why standard statistical methods will fail in the estimation of the causal effect of such exposures.

1.2.4 Time varying confounding

A patient admitted to the ICU can acquire infection on each day during his stay and the time of infection is evidently not known upon admission. For each patient, we therefore observe A_t which indicates 1 for infection at or prior to time t and 0 otherwise. Here t denotes days since ICU admission. Let $\overline{A}_t = (A_0, A_1, \ldots, A_t)$ be the infection history through day t. L_0 is a collection of baseline confounders and L_t time-dependent severity of illness indicators. Similarly, $\overline{L}_t = (L_0, L_1, \ldots, L_t)$ denotes the severity of illness history through day t. On each day t we assume that all confounders for the relation between infection and mortality are measured by $\overline{L}_t$ and $\overline{A}_{t-1}$. Actually, as in any observational study, it is not possible to check (test) from the observed data on L_t, A_t and Y whether there is confounding by unmeasured risk factors U_t. We only can hope that confounding by U_t is very weak.

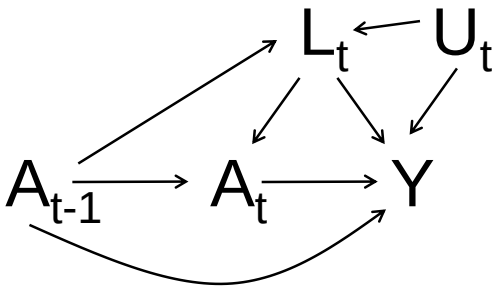

Figure 1.3: Causal DAG to illustrate the principle of time varying confounding

The key problem of time dependent confounding is illustrated in the simplified DAG given in Figure 1.3. Because a patient's health status L_t is affected by prior infection A_{t-1}, several of these covariates are *intermediate on the causal path* from early infection to mortality, and therefore should not be adjusted for. This is not only because standard regression would remove indirect infection effects mediated through severity of illness, but also because it would induce a so-called collider-stratification bias (Greenland, 2003). On the other hand L_t is a *confounder*

for the relation between late infection A_t and mortality and therefore should be adjusted for. A solution to this problem is given by the use of *marginal structural models* MSMs (Hernán et al., 2000).

1.2.5 Marginal structural models (MSMs)

Marginal structural models (MSMs) are models for the marginal distribution of a certain couterfactual outcome and were developed by Robins et al. (2000). They enable proper adjustment for time-dependent confounding. Let $Y(\overline{a})$ be the mortality status Y that would have been observed at the end of follow-up had *all* subjects followed the hypothetical infection path $\overline{a}$. Such potential infection path $\overline{a} = (a_1, a_2, ..., a_E)$ equals a vector of elements a_t, for t going from day 1 (i.e. admission) in the ICU to a fixed end-of-follow-up time E, representing a hypothetical regime in which *all* ICU patients would be uninfected in the ICU during all days t with $a_t = 0$ and infected in the ICU on all remaining days t' with $a_{t'} = 1$.

As in Robins et al. (2000), the logistic MSM for $Y(\overline{a})$ might take the form

$$\text{logit}P[Y(\overline{a}) = 1] = \beta_0 + \beta_1 cum(\overline{a}) \tag{1.1}$$

with $cum(\overline{a}) \equiv \sum_{t=1}^{t-1} a_t$ denoting the number of days a patient would have been infected prior to time t under infection path $\overline{a}$. The parameter e^{β_1} expresses the effect of an extra day of infection on the odds of death, had *all* patients had the same infection history $\overline{a}$. If the exposure-outcome relationship is not confounded the parameters of model (1.1) can be unbiasedly estimated by a logistic regression model of the observed data

$$\text{logit}P[Y = 1|\overline{A} = \overline{a}] = \beta_0' + \beta_1' cum(\overline{a}) \tag{1.2}$$

If the exposure-outcome relationship is indeed confounded by $\overline{L}_t$ and $\overline{A}_{t-1}$, then $\beta_1' \neq \beta_1$ and the standard logistic regression estimate of the causal effect will be biased. If all confounders

for the exposure-outcome relationship are measured by $\overline{L}_t$ and $\overline{A}_{t-1}$, then the causal parameter β_1 can be estimated using inverse probability weighting (IPW) to correct for confounding. Using the observational data, we estimate the exposure effect from an outcome regression model but weigh each observation by the inverse of the probability of the observed exposure level given the time dependent confounders.

A detailed theoretical description of this approach can be found in **chapter 2**, a more heuristic explanation of how marginal structural models are fitted is given in **chapter 3** of this chapter.

1.3 Competing risk analysis

In this section we will provide an introduction to competing risk analysis. A very good reference on this topic is Chapter 8 in Kalbfleisch and Prentice (2002); a simpler introduction is the book by Pintilie (2006). There is also a nice tutorial by Putter et al. (2007). The main goal of this section is to explain the two approaches to competing risk analysis (cause specific hazard vs. sub-distribution hazard) and give some arguments why we prefer one approach over the other when the interest is in estimating the attributable mortality of nosocomial infections. A more technical explanation how a causal analysis of competing events is done is given in **chapter 2**.

1.3.1 Competing events in the ICU

Not for every patient admitted to the ICU we observe the time of death at the end of follow up because the survival times are *censored* by ICU-discharge. The Cox model assumes that censoring is *non-informative* or that at each time, among patients who are alive and still in the ICU at that time, the probability of discharge is independent of the actual survival time. A major problem with the use of survival methods to analyze ICU mortality lies in the fact that patients who get discharged from the ICU on a given day since admission are not comparable with those who stay in the ICU on that day. They are discharged from the ICU because they need no more

intensive care due to improved health conditions and therefore have a lower mortality risk than the average. In that case, censoring carries information about, or depends on the survival time.

In the analysis of ICU-data we will therefore consider ICU death as the event of interest and discharge from the ICU as the competing event because it prevents the event of interest from occurring (once a patient is discharged form the ICU cannot die in the ICU), rather than just preventing us from seeing it happen (censoring).

1.3.2 Quantifying the attributable mortality

A variety of measures (risk differences, odds ratios, relative risks, hazard ratios,...) to quantify the attributable mortality of infections are used in the ICU-literature (Melsen et al., 2009). Some of those measures e.g. hazard ratios, are difficult to assign a causal interpretation because they have a sort of built in selection bias: the risk sets (the number of subjects at risk of experiencing the event of interest) may not stay comparable over time (Hernán, 2010). This is even so in the context of randomized experiments where the 2 treatment arms, while comparable at the onset of the study, may not stay comparable over time when mortality is differential between both arms.

For each patient, we observe the time T from ICU admission to death or ICU discharge, whichever comes first and the event type ϵ, which indicates 1 for the event of interest (ICU death) and 2 for the competing event (ICU discharge). In **chapter 2**, we infer the impact of acquiring infection in the ICU on the *Cumulative Incidence Function (CIF)* of ICU-death ($\epsilon = 1$), which is the probability of dying within the ICU before a given time, as a function of time or $F_1(t) = P(T \leq t, \epsilon = 1)$. The CIF was then used to estimate the attributable mortality of infection as the *population attributable fraction* (PAF) of ICU-mortality related to infection. Generally, the PAF is defined (Rothman and Greenland, 1998) (page 295) as *the fraction of all cases (exposed and unexposed) that would not have occurred if exposure had not occurred.* On

each day, we calculated the PAF as the difference between the observed ICU-mortality ($F_1(t)$) and the ICU-mortality that would have been observed (counterfactual) for the same population if infection were prevented for all ($F_{\bar{0}1}(t)$), divided by the observed ICU-mortality (see **chapter 2** for more details).

$$\frac{F_1(t) - F_{\bar{0}1}(t)}{F_1(t)}, \tag{1.3}$$

Under monotonicity, it can be interpreted as the percentage of the observed ICU-deaths that could be avoided by preventing infection, or, as the percentage of the observed ICU-deaths who died because of infection (Tian and Pearl, 2000), which provides a direct answer to tho the question how many patients died *with* or *from* infection. The monotonicity assumption is in epidemiology, often expressed as *no prevention*, in which no individual in the population can be helped by exposure to the risk factor.

We are not the first ones to use the concept of the population attributable fraction to estimate the attributable mortality of nosocomial infections. Schumacher et al. (2007) used the following definition:

$$PAF(t) = \frac{F_1(t) - F_1(t|A_t = 0)}{F_1(t)}$$

where $F_1(t)$ equals the observed mortality before time t and $F_1(t|A_t = 0)$ the mortality before time t within the subgroup of patients who were not infected before time t. These measures were calculated using a multi state model approach (cause specific hazard analysis). Unlike our proposed measure (1.3) this definition involves comparing different subsets of the patients who are still in the risk set at each time, and thus cannot be assigned a causal interpretation.

1.3.3 Cause specific vs. subdistribution hazard

We can consider competing risk data as arising from a multistate model (Andersen et al., 2002; Beyersmann et al., 2006; Schumacher et al., 2007) shown in Figure 1.4.

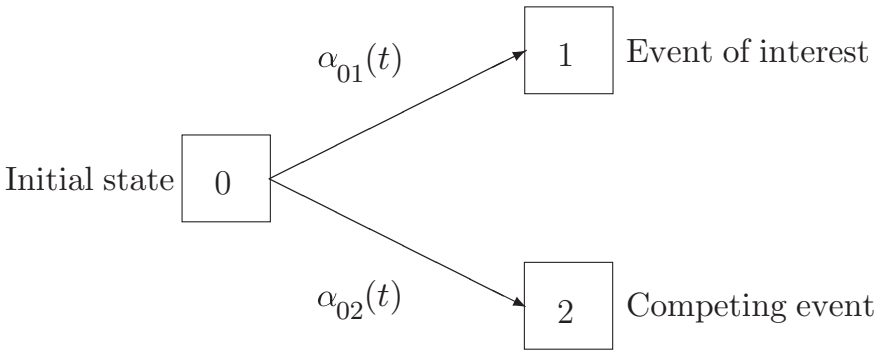

Figure 1.4: Competing risk model as a multistate model

The state which an individual is in at a given point in time is denoted by $X_t \in \{0, 1, 2\}$. A patient is admitted to the ICU and stays in state 0. When a patient dies he/she moves to state 1 (event of interest), after discharge a patient moves to state 2 (competing event). Individuals are moving out of the initial state 0 at their failure time T (time to death or ICU discharge) and their cause of failure is then equal to the state X_T where they moved to. The whole process is determined through the transition intensities α_{0k} with k = 1 or 2 which are the usual *cause specific hazards*. For an event of type k we can define the *cause-specific hazard function* as the hazard of failing from cause k in the presence of the competing events

$$\alpha_k(t) = \lim_{dt \to 0} \frac{1}{dt} P(t \leq T \leq t + dt, \epsilon = k | T \geq t)$$

For this type of hazard it is possible to define a proportional hazards which is of the form

$$\alpha_k(t; X(t)) = \alpha_{k0}(t) \exp(b_k X(t))$$

where $\alpha_{k0}(t)$ is a non specified baseline hazard function and b_k is the unknown regression parameter. For the estimation of the cause specific hazard function of ICU-death, the standard Cox proportional hazards model can be employed. The time to ICU-death is then treated as the

observed survival time and the time to ICU-discharge is treated as additional censoring times. ICU-discharge and censoring is treated in an identical manner because both just imply a reduction of the number of subjects at risk.

As mentioned in the previous section, our primary interest lies in the estimation of the attributable mortality of infection as a function of the CIF. There is no one-to-one relationship between the cause specific hazards and the CIF. In order to use the cause specific hazards approach to make inference on the CIF it is necessary to estimate the regression parameters for the proportional hazards models for $\alpha_{01}(t)$, $\alpha_{02}(t)$ together with the baseline hazards. A more direct approach is to model the *subdistribution hazard* which is defined as

$$\lambda_k(t) = \lim_{dt \to 0} \frac{1}{dt} P(t \leq T \leq t + dt, \epsilon = k | T \geq t \text{ or } (T < t \text{ and } \epsilon \neq k))$$

Fine and Gray (1999) proposed proportional hazards models of the form

$$\lambda_k(t; X(t)) = \lambda_{k0}(t) \exp(\beta_k X(t))$$

to model the subdistribution hazard $\lambda_k(t; X(t))$, where $\lambda_{k0}(t)$ is a non specified baseline hazard function and β_k is the regression parameter. By definition, the subdistribution hazard is directly related to the CIF of failure from cause k, by $\lambda_k(t) = -dlog\{1 - F_k(t)\}/dt$. This one-to-one relationship was our main motivation to develop a method to adjust for time varying confounding in the *subdistribution hazard analysis* of competing risks because, due to the difficult causal interpretation of hazard ratios, we defined the attributable mortality of infection on the basis of the CIF.

In general, the main difference between the two approaches lies in the definition of the risk set. Like in the Cox model for the cause specific hazard, the regression model for the subdistribution hazard includes individuals who have not failed from any cause by time t (risk set at time

t defined by $j : t \leq T$), but additionally those who have previously failed from the competing cause before t. The subdistribution hazard represents the probability of observing the event of interest in the next time interval while knowing that either the event of interest did not happen until then or that the competing risks event was observed. Subjects for whom the competing event is observed stay in the risk set at all times.

2

Adjusting for time-varying confounding in the subdistribution analysis of competing risks

This chapter is adapted from *Bekaert et al. (2010b)* and published in a special issue of *Lifetime data analysis* on survival analysis, longitudinal analysis and causal inference. The original coauthors were my promotor, Stijn Vansteelandt and Karl Mertens who provided the data for analysis.

Summary

The assessment of the attributable mortality due to nosocomial infections in the intensive care

unit (ICU) remains controversial, with several studies describing effect estimates ranging from being neutral to extremely risk increasing. Interpretation of study results is further hindered by inappropriate adjustment (a) for censoring of the survival time by discharge from the ICU, and (b) for time-dependent confounders on the causal path from infection to mortality. Previously (Vansteelandt et al., 2009), this was accommodated through inverse probability of treatment and censoring weighting. Because censoring due to discharge from the ICU is so intimately connected with a patient's health condition, the ensuing inverse weighting analyses suffer from influential weights and rely heavily on the assumption that one has measured all common risk factors of ICU discharge and mortality. In this paper, we consider ICU discharge as a competing risk in the sense that we aim to infer the risk of 'ICU mortality' over time that would be observed if nosocomial infections could be prevented for the entire study population. For this purpose we develop marginal structural subdistribution hazard models with accompanying estimation methods. In contrast to subdistribution hazard models with time-varying covariates, the proposed approach (a) can accommodate high-dimensional confounders, (b) avoids regression adjustment for post-infection measurements and thereby so-called collider-stratification bias, and (c) results in a well-defined model for the cumulative incidence function. The methods are used to quantify the causal effect of nosocomial pneumonia on ICU mortality using data from the National Surveillance Study of Nosocomial Infections in ICU's (Belgium).

2.1 Introduction

Nosocomial (i.e., hospital-acquired) infections are highly prevalent in intensive care unit (ICU) patients because of their poor health conditions and, additionally, because of the prevalent use of invasive treatments (e.g., mechanical ventilation) that make it more difficult for the human body to conquer hostile bacteria. They form a major public health problem in the Western world. For instance, in Belgium (which has a population size of about 10.5 million people), it is roughly estimated that nosocomial infections contribute to the mortality of 2600 patients per year, to

700 000 extra hospital days and to a total annual cost of 400 million EUR (Vrijens et al., 2009). There is a longstanding interest in precise knowledge of the effect of these infections on mortality (Safdar et al., 2005; Davis, 2006; Heyland et al., 1999). This stems from the desire to better understand their severity as well as the importance of a timely treatment. This is additionally motivated by the need to justify the high costs that these infections impose on the society as a result of prolonged hospital stay and extra hospitalisation costs.

The question whether nosocomial infections are severely harmful, remains a subtle one for several reasons. First, infection must be viewed in combination with antibiotic treatment, which has beneficial effects if applied appropriately (Kollef, 2003) but may have harmful effects otherwise (e.g., due to antibiotic resistance). Second, patients who acquire these infections and subsequently die, may have been so severely ill that they would have died even without the infection. In view of this, it will be crucial to have the ability to adjust for the evolution in disease severity over time.

Despite the wide literature on the effect on mortality of acquiring a nosocomial infection in the ICU, assessment of this effect remains controversial, with several studies describing effect estimates ranging from being neutral to extremely risk increasing (Heyland et al., 1999; Fagon et al., 1993; Papazian et al., 1996; Timsit et al., 1996a; Girou et al., 1998; Bercault and Boulain, 2001). Interpretation of study results is further hindered by time dependent bias (i.e., bias due to ignoring the time-dependent nature of the data (Beyersmann et al., 2008; Van Walraven et al., 2004)) and inappropriate adjustment for time-dependent confounding and for informative censoring due to discharge from the ICU (Vansteelandt et al., 2009). Informative censoring (Kalbfleisch and Prentice, 2002) results from the decision to discharge a patient from the ICU being closely related to his/her health status. Patients who get discharged from the ICU on a given day since admission are therefore not comparable with those who stay in the ICU on that day. If a rich collection of common risk factors of mortality and ICU discharge were measured, one could in

principle accommodate this using inverse probability of censoring weighted estimators (Vansteelandt et al., 2009; Robins and Rotnitzky, 1992; Keiding et al., 1999). However, the resulting analyses rely heavily on the assumption that all such risk factors have been measured and suffer from influential weights because censoring tends to be strongly associated with mortality in this setting (Vansteelandt et al., 2009). In this paper, we will therefore shift the focus and estimate the impact of acquiring infection in the ICU on the Cumulative Incidence Function (CIF) of ICU death (Prentice et al., 1978; Fine and Gray, 1999; Pepe and Mori, 1993). This is the probability of dying within the ICU before a given time, as a function of time. Focussing on this estimand not only rules out the censoring problem (since ICU death by a given time is observable for all patients so that censoring becomes purely administrative due to end of follow-up); ICU mortality is also of primary interest to ICU physicians, is frequently used in this domain (see e.g. Resche-Rigon et al. (2006a); Valls et al. (2007); Wolkewitz et al. (2009)) and would be representative of overall mortality if the hazard of death were relatively small after ICU discharge. A similar focus is taken by e.g. Wolkewitz et al. (2009, 2008) and Beyersmann et al. (2006) who consider discharge from the ICU and ICU mortality as competing risks to assess the impact of infection on ICU mortality and length of stay in the ICU.

Given our focus on the CIF of ICU death, a natural approach would be to consider the infection status as a time-dependent covariate in a model for the subdistribution hazard (i.e. the hazard function corresponding to the CIF) (Wolkewitz et al., 2008; Beyersmann and Schumacher, 2008; Barnett and Graves, 2008). This approach acknowledges that patients need to survive and stay within the ICU long enough before acquiring infections (Beyersmann et al., 2008; Van Walraven et al., 2004). However, because patients who do and do not acquire infection at a given time are not comparable in terms of severity of illness, observed mortality differences between these groups cannot be fully attributed to an infection effect. Including measures of disease severity as additional time-varying covariates in the model for the subdistribution hazard does not accommodate this. First, it inevitably raises difficulties as to how to define these covariates, which were

only measured during the ICU stay, after discharge (i.e., when the competing event has occurred) (Beyersmann and Schumacher, 2008; Latouche et al., 2005). Second, because a patient's health status is affected by prior infection, several of these covariates are intermediate on the causal path from early infection to mortality, and therefore should not be adjusted for via standard regression methods (Hernán et al., 2000). This is not only because standard regression would remove indirect infection effects mediated through severity of illness, but also because it would induce a so-called collider-stratification bias (Greenland, 2003). Third, subdistribution hazards adjusted for time-dependent covariates do not result in a model for the CIF anymore (Beyersmann and Schumacher, 2008), just like standard survival Cox models with time-dependent covariates yield no model for the survival function (Andersen et al., 1993). This complicates the interpretation of the infection effect by the fact that a hazard ratio, unlike a comparison of the CIFs, does not lend itself to a simple causal interpretation. This is so because, even if infected and uninfected patients were exchangeable (Hernán and Robins, 2006) upon ICU admission, the risk sets of surviving infected and uninfected patients may not stay exchangeable over time (Hernán et al., 2004).

In view of these limitations, we will develop marginal structural models (Robins et al., 2000) for the counterfactual subdistribution hazard with accompanying estimation methods. Using these models, we will be able to infer the risk of ICU death over time that would be observed if nosocomial infections could be prevented for the entire study population. The proposed framework for causal effect estimation (a) can accommodate high-dimensional confounders, (b) avoids regression adjustment for post-infection measurements, and (c) results in a model for the CIF. The methodology is developed for complete data, but may extended to right-censored data using additional inverse probability of censoring weights (Robins and Rotnitzky, 1992; Fine and Gray, 1999).

This chapter is organized as follows. In Section 2.2 we start by introducing the main concepts

and develop inference under a model which imposes no parametric restrictions on the counterfactual cumulative incidence function. This is then extended in Section 2.4 to semi-parametric discrete-time marginal structural models for the subdistribution hazard. The methods are used in Sections 2.3.3 and 2.4.2 to quantify the effect of nosocomial pneumonia on ICU mortality using data from the National Surveillance Study of Nosocomial Infections in ICU's (Belgium). We end in Section 2.5 with a discussion on the assumptions and the relative advantages and limitations of the proposed methodology.

2.2 Notation, definitions and assumptions

2.2.1 Observed data

For each patient, we observe the time T from ICU admission to death or ICU discharge, whichever comes first and the event type ϵ, which indicates 1 for the event of interest (ICU death) and 2 for the competing event (ICU discharge). In addition, for each patient, we observe a discrete-time counting process $\{A_t, 0 \leq t \leq T\}$ until death or ICU discharge, whichever comes first, where t denotes days since ICU admission. Here, A_t indicates 1 for ICU acquired infection at or prior to time t and 0 otherwise. Finally, for each patient, we observe a collection of baseline confounders L_0 and time-dependent confounders L_t on each day t, which will be listed in Section 2.3.3. Our interest focuses on the Cumulative Incidence Function (CIF) of ICU death: $P\left(T \leq t, \epsilon = 1\right)$.

2.2.2 Counterfactual subdistribution hazard and cumulative incidence function

Unlike a competing risk analysis with time-dependent infection status (Wolkewitz et al., 2008; Beyersmann and Schumacher, 2008), we will evaluate the CIF for the entire population under different *hypothetical or potential infection paths* to guarantee that we compare the same patient population each time. This requires that we define for each patient a counterfactual event time $T_{\overline{a}}$ (Rubin, 1978; Robins, 1986), which represents the time until ICU death or ICU discharge, whichever comes first, which an ICU patient would - possibly contrary to fact - have had under

the given hypothetical infection path $\overline{a}$. Such potential infection path $\overline{a} = (a_1, a_2, ..., a_E)$ equals a vector of elements a_t, for t going from day 1 (i.e. admission) in the ICU to a fixed end-of-follow-up time E, representing a hypothetical regime in which *all* ICU patients would be uninfected in the ICU during all days t with $a_t = 0$ and infected in the ICU on all remaining days t' with $a_{t'} = 1$. Correspondingly, we define $\{\epsilon_{\overline{a}}(t), t \geq 0\}$ to be a discrete-time stochastic process expressing the counterfactual event status under infection path $\overline{a}$ on each day t. This indicates 1 (or 2) on day t if the patient would have died in (been discharged from) the ICU under that potential infection path, and 0 if no event would have occurred by that day. In addition, we define $\epsilon_{\overline{a}} \equiv \epsilon_{\overline{a}}(T_{\overline{a}})$. For an event of type k ($k = 1, 2$), we thus conceptualize the corresponding counterfactual subdistribution event time

$$\theta_{\overline{a}k} \equiv \inf\{t \in [0, \infty[\colon \epsilon_{\overline{a}}(t) = k\}.$$

For $k = 1$ ($k = 2$), this is equal to $T_{\overline{a}} = \inf\{t \in [0, \infty[\colon \epsilon_{\overline{a}}(t) \neq 0\}$ for patients who die within (get discharged from) the ICU under the given infection path $\overline{a}$ and ∞ for patients who get discharged from (die within) the ICU under that infection path. We further define the counterfactual CIF corresponding to event type k as

$$F_{\overline{a}k}(t) \equiv P\left(T_{\overline{a}} \leq t, \epsilon_{\overline{a}} = k\right) = P\left(\theta_{\overline{a}k} \leq t\right),$$

for all $t \geq 0$. This is the probability that, under infection path $\overline{a}$, an event of type k occurs at or before time t.

Our focus in this article will be on inferring the CIF under monotone infection paths in which $a_t \geq a_s$ for $t > s$. This will enable us to infer the effect of acquiring infection on a given day, disregarding the possible impact of the duration of infection. For instance, we can define the causal effect of acquiring infection on day t since admission in the ICU on the cumulative incidence of ICU death as a comparison of 2 counterfactual CIFs under different monotone

infection paths $(a_1, ..., a_{t-1}, a_t, ..., a_E) = (0, ..., 0, 1, ..., 1)$ and $(a_1, ..., a_E) = (0, ..., 0)$, indicating regimes where *all* patients acquire infection on day t since admission versus never in the ICU, respectively. By considering monotone infection paths only, we will be able to ignore information on the duration of infection, which is often imprecisely measured, and will thus average over patients with different duration of infection. Alternatively, the causal effect of infection can be reported in terms of the population attributable fraction of ICU mortality related to infection, which we define at each time t as

$$\frac{\hat{F}_1(t) - \hat{F}_{\overline{01}}(t)}{\hat{F}_1(t)}, \tag{2.1}$$

where $\hat{F}_1(t)$ is the estimated observed CIF of ICU mortality at time t and $\hat{F}_{\overline{01}}(t)$ is the estimated counterfactual CIF of ICU mortality at time t under the no infection path. This expresses what percentage of the observed ICU deaths by time t could be avoided by preventing infection. A related comparison of observed and counterfactual infection-free responses was considered in (Schulgen and Schumacher, 1996) to estimate the prolongation of stay attributable to nosocomial infections. We believe that (2.1) links more closely to the originally concept of attributable risk than other proposed measures (Schumacher et al., 2007), in particular because at each time it compares the same population twice. In the appendix we give the derivation of the variance and 95% confindence interval of (2.1) .

2.2.3 Assumptions

To be able to identify the counterfactual CIF under all infection paths, the following 2 assumptions must be made (see Section 2.5 for a further discussion of the assumptions). First, we will make the so-called consistency assumption which links the counterfactual data to the observed data by assuming that $T_{\overline{a}}$ and $\epsilon_{\overline{a}}$ equal the observed event time T and status ϵ for patients whose observed infection data are *compatible with* (i.e. would have been observed under) the given infection path $\overline{a}$ at that time (see also Section 2.3.1). To appreciate this assumption, note that

our definition of 'infection path' assumes the existence of an intervention whereby infection is prevented for all patients during a given period of time and acquired at a given later time t (provided they are still in the ICU). The consistency assumption then states that this intervention is non-invasive in the sense that it has no effect other than causing infection at and not before time t (VanderWeele and Vansteelandt, 2009). In particular, it assumes that for patients who would naturally acquire (prevent) infection on day t, the same event time and status would be observed under that intervention as was naturally observed. In this sense, the population attributable fraction (2.1) reflects a 'pure' infection effect that would be realized if patients would somehow 'naturally' avoid infections. Second, we will make the no unmeasured confounders assumption that, for each time t, among uninfected patients who have been in the ICU with the same covariate history $\overline{L}_{t-1} \equiv (L_0, ..., L_{t-1})$ up to time $t-1$, those who acquire infection at time t are exchangeable with those who do not in the sense that $\underline{\epsilon_{\overline{a}}(t)} \coprod A_t | \overline{A}_{t-1}, \epsilon_{\overline{a}}(t-1) = 0, \overline{L}_{t-1}$ for all t and all infections paths $\overline{a}$; here, for any vector (W_t), $\overline{W}_t$ refers to the history $(W_1, ..., W_t)$ and $\underline{W}_t$ refers to the future $(W_t, ..., W_E)$. Here we assume that, as was the case in our data analysis (see Section 2.3.3), the infection status was measured on day t for all patients who did not die or get discharged before day t.

2.3 Estimation under a nonparametric model for the counterfactual cumulative incidence function

2.3.1 Compatible infection paths

To infer the counterfactual CIF under a given infection path $\overline{a}$, we must select those patients whose observed data are compatible with it: i.e. those for whom we would have observed the same event time and status had such infection path been implemented. As an illustration, we will show how this can be done for the infection path $\overline{a}(0) = (0, 0, 0, ..., 0)$ under which patients stay infection-free and for the infection path $\overline{a}(5) = (0, 0, 0, 0, 1, ..., 1)$ under which patients aquire infection 5 days after admission. Consider the following 4 hypothetical patients

in Figure 2.1. Patient 1 stays in the ICU without infection for at least 10 days, patients 2 and 3 acquire infection in the ICU on day 5 and 4 respectively, and finally patient 3 gets discharged from the ICU without infection on day 3. The days on which a patient is compatible with the hypothetical infection paths $\overline{a}(0)$ and $\overline{a}(5)$ are indicated with a larger font size in Figures 2.1 A and B, respectively, as we now explain.

No infection Figure 2.1 (A) illustrates that patients who have not acquired infection and are still in the ICU by time t are compatible with $\overline{a}(0)$ at time t. For instance, at $t = 3$, all patients in Figure 2.1 (A) are compatible with $\overline{a}(0)$; at $t = 4$, patients 1 and 2, unlike patient 3, are compatible with $\overline{a}(0)$. In addition, patients who get discharged or die without infection before time t (e.g. patient 4) are compatible with $\overline{a}(0)$ at time t because their same observed data would have been obtained under a regime in which infection is prevented for all patients.

Acquiring infection on day 5 Figure 2.1 (B) illustrates that on all days $t < 5$, this infection path coincides with the no infection path and thus the same patients are compatible with both $\overline{a}(5)$ and $\overline{a}(0)$ on those days. For instance, patient 1 in Figure 2.1 (B) is compatible with $\overline{a}(5)$ on all days $t < 5$; patient 3 is compatible with $\overline{a}(5)$ until the day before acquiring infection and obviously, patients who acquire infection on day 5 (i.e. patient 2) are compatible with $\overline{a}(5)$ on all days t. In addition, patients who get discharged or die without infection before day 5 (i.e. patient 4), are compatible with $\overline{a}(5)$ on day t because they would also have been discharged/have died prior to day 5 under a regime in which infection is acquired on day 5 (and not before) by all patients who are still in the ICU on that day.

Note from the previous examples that a patient's observed data can be compatible with (and thus carry information about) multiple infection paths. For instance, a patient who is discharged without infection 20 days after admission is compatible with all infection paths in which infection is acquired after day 20. When the end-of-follow-up time E is 30 days, this patient thus has 10 compatible infection paths (see Figure 2.2). In general, a patient who gets discharged without infection at a given time t is compatible with all infection paths $\overline{a}(t + s)$ for $s > 0$.

	t	1	2	3	4	5	6	7	8	9	10	...
	$a_t(0)$	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**...**
Patient 1	A_t	0	0	0	0	0	0	0	0	0	0	...
Patient 2	A_t	0	0	0	0	1	1	1	1	1	1	...
Patient 3	A_t	0	0	0	1	1	1	1	1	1	1	...
Patient 4	A_t	0	0	0								

A

	t	1	2	3	4	5	6	7	8	9	10	...
	$a_t(5)$	**0**	**0**	**0**	**0**	**1**	**1**	**1**	**1**	**1**	**1**	**...**
Patient 1	A_t	0	0	0	0	0	0	0	0	0	0	...
Patient 2	A_t	0	0	0	0	1	1	1	1	1	1	...
Patient 3	A_t	0	0	0	1	1	1	1	1	1	1	...
Patient 4	A_t	0	0	0								

B

Figure 2.1: Selection of patients compatible with infection paths $\overline{a}(0)$ (A) and $\overline{a}(5)$ (B).

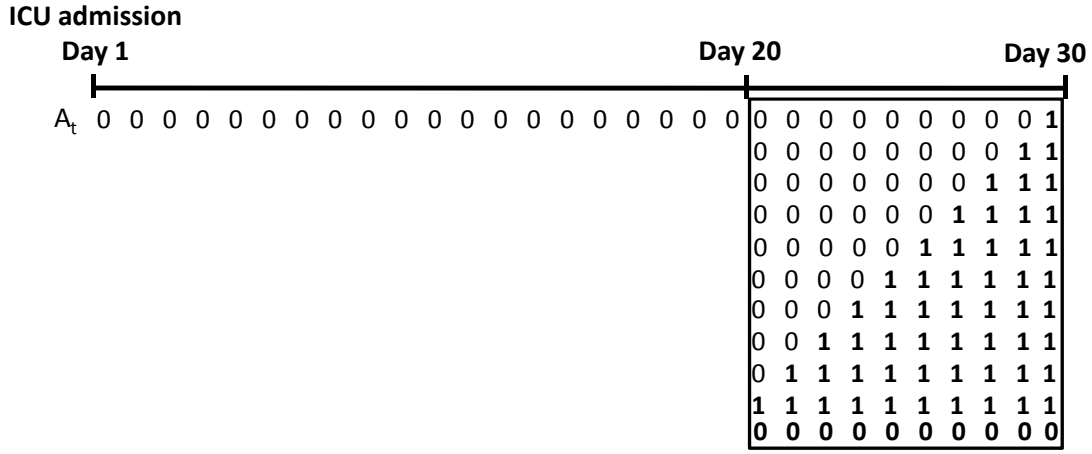

Figure 2.2: Compatible infection paths for a patient who gets discharged without infection 20 days after admission (endpoint is 30 day ICU mortality).

2.3.2 Estimation under a given infection path

In this section, we discuss the estimation of the counterfactual CIF for a given infection path $\overline{a}$. Define $d_{\overline{a}kt}$ to be the number of events of type k that occur at time t for the $n_{\overline{a}t}$ patients who are compatible with infection path $\overline{a}$ at that time. If there were no confounding, then the counterfactual cumulative incidence at time t for event type k under infection path $\overline{a}$ could be estimated as the percentage of observed events of type k at or before time t amongst patients who are compatible with $\overline{a}$ at that time:

$$\frac{\sum_{s=1}^{t} d_{\overline{a}ks}}{n_{\overline{a}t}} = \frac{\sum_{i=1}^{n} I\left(\epsilon_i(t)=k\right) \prod_{s=1}^{t} \left\{I(A_{is}=a_s) I\left(\epsilon_i(s)=0\right) + I\left(\epsilon_i(s)\neq 0\right)\right\}}{\sum_{i=1}^{n} \prod_{s=1}^{t} \left\{I(A_{is}=a_s) I\left(\epsilon_i(s)=0\right) + I\left(\epsilon_i(s)\neq 0\right)\right\}}. \quad (2.2)$$

In general, this is a biased estimator of the counterfactual CIF at time t because it uses only data from people whose observed infection status was compatible with the chosen infection path at that time, and these may not form a random subset of the target population. We will therefore refer to it as the *naïve* estimator of the counterfactual CIF. To adjust for this bias, we will use inverse probability of treatment weighting (IPTW) and estimate the counterfactual cumulative incidence at time t for event type k under infection path $\overline{a}$ as:

$$\hat{F}_{\overline{a}k}(t) = \frac{\sum_{i=1}^{n} I\left(\epsilon_i(t)=k\right) W_{i\overline{a}t}}{\sum_{i=1}^{n} W_{i\overline{a}t}}, \quad (2.3)$$

where

$$W_{i\overline{a}t} \equiv \prod_{s=1}^{t} \frac{I(A_{is}=a_s) I\left(\epsilon_i(s)=0\right) + I\left(\epsilon_i(s)\neq 0\right)}{P(A_{is}=a_s|\epsilon_i(s), \overline{A}_{i,s-1}, \overline{L}_{i,s-1})}. \quad (2.4)$$

Here, $P(A_{is}=a_s|\epsilon_i(s), \overline{A}_{i,s-1}, \overline{L}_{i,s-1})$ denotes the probability of having infection status a_s at time s conditional on the observed covariates $(\epsilon(s), \overline{A}_{s-1}, \overline{L}_{s-1})$. Because patients who are discharged or have died before time s stay compatible with infection path $\overline{a}$ when they were compatible with this path at the time of discharge/death, no further selection is made within that subgroup and thus we define $P(A_{is}=a_s|\epsilon_i(s), \overline{A}_{i,s-1}, \overline{L}_{i,s-1})$ to equal 1 when $\epsilon(s) \neq 0$. When $\epsilon(s)=0$, this probability can be estimated using a pooled logistic regression model.

In the Appendix, we show that (2.3) is a consistent estimator of the counterfactual cumulative incidence at time t under infection path $\overline{a}$ when, in addition to the assumptions of Section 2.2.2, $P(A_t = a_t|\epsilon(t) = 0, \overline{A}_{t-1}, \overline{L}_{t-1})$ is correctly specified at each time t. In the Appendix, we further show that for given t, a conservative, asymptotic $(1-\alpha)100\%$ confidence interval for $F_{\overline{a}k}(t)$ can be obtained as

$$\text{expit}\left[\text{logit}\left\{\hat{F}_{\overline{a}k}(t)\right\} \pm z_{\alpha/2}\sqrt{\frac{1}{n}\frac{\widehat{\text{Var}}(U_{it\overline{a}k})}{\hat{F}_{\overline{a}k}(t)\{1-\hat{F}_{\overline{a}k}(t)\}}}\right], \tag{2.5}$$

where $z_{\alpha/2}$ is the $(1-\alpha/2)100\%$ percentile of the standard normal distribution, $U_{it\overline{a}k}$ is the weighted residual

$$U_{it\overline{a}k} \equiv W_{i\overline{a}t}\left\{I\left(\epsilon_i(t) = k\right) - \hat{F}_{\overline{a}k}(t)\right\} \tag{2.6}$$

and $\widehat{\text{Var}}(U_{it\overline{a}k})$ refers to the sample variance of $U_{it\overline{a}k}$.

2.3.3 Data analysis

We use the proposed techniques for the analysis of the National ICU Surveillance Study in Belgium (Vansteelandt et al., 2009; Suetens et al., 1999, 2003). A step by step explanation together with detailed R-code is provided in the appendix. All ICU's in Belgian hospitals were invited to participate in this surveillance study on a voluntary basis. For all patients admitted to the ICU, data were recorded on personal characteristics, reasons for ICU admission, baseline health status, and daily indicators of received invasive treatments and acquired infections in the ICU. Nosocomial pneumonia (NP) was defined as pneumonia acquired by patients after the second day of ICU stay, to exclude infections that were in incubation upon enrollment in the ICU. The third day of stay in ICU will therefore be the starting point for our analysis, thus excluding patients who stayed less than 3 days. We restrict the analysis to surveillance data collected for the years 2002 and 2003 in three of the largest hospitals, which have accurate daily measurements

of received invasive treatments and acquired infections. A total of 4288 ICU patients were analysed. Of the 360 (8.4%) patients who acquired NP in the ICU and stayed more than 2 days, 75 (20.8%) died in the ICU, as compared to 360 (9.2%) deaths among the 3928 patients who remained NP-free in the ICU; of these 360 (infected) patients, 245 or 68.2% (328 or 91.3%) acquired infection during the first week (first two weeks) after admission. Among patients who stayed more than 2 days in the ICU, the median length of stay in the ICU was 4 days (IQR 3, 95th percentile 14) for those without an infection history and 18 days (IQR 19, 95th percentile 30) for the remaining patients.

To estimate the CIF of ICU death under the no infection path $\overline{a}(0)$, we select those patients for whom we would have observed the same event time and status had such infection path been implemented (see Figure 2.1 (A) for an illustration). The number of patients $n_{\overline{a}(0)t}$ whose observed data are compatible with $\overline{a}(0)$ at each time t, along with the number $d_{\overline{a}(0)}$ of those who died at or before time t are given in the second and third column of Table 2.1. Similar numbers are reported in columns 4 and 5 for the infection path $\overline{a}(5)$. Note that $n_{\overline{a}(0)t}$ diminishes each time with the number of patients who acquire infection at that time. Patients who get discharged without infection by a given time t stay compatible with this path until the end-of-study time E, which was chosen to equal 30 days in our analyses. Note also that, by construction, identical results are obtained for the infection paths $\overline{a}(0)$ and $\overline{a}(5)$ up to day 5. From day 5 onwards, the number of patients whose data are compatible with $\overline{a}(5)$ stays constant until the end-of-study time as it equals the number of patients who acquire infection on day 5 together with those who died or were discharged without infection prior to day 5.

The naïve estimator (2.2) corresponding to $\overline{a}(0)$ is obtained as the ratio of $d_{\overline{a}(0)}$ over $n_{\overline{a}(0)}$. The resulting estimates are displayed in Figure 2.3 (A) and suggest roughly a 1% absolute reduction in ICU mortality at 30 days if infection were avoided for all patients. These estimates are biased because they ignore confounding. We therefore calculate the IPTW-estimator (2.3). This requires estimating the probability of acquiring infection at each time t. We have built a

	Naïve				IPTW			
t	$n_{\overline{a}(0)}$	$d_{\overline{a}(0)}$	$n_{\overline{a}(5)}$	$d_{\overline{a}(5)}$	$n_{\overline{a}(0)}$	$d_{\overline{a}(0)}$	$n_{\overline{a}(5)}$	$d_{\overline{a}(5)}$
3	4228	86	4228	86	4290	87	4290	88
4	4167	144	4167	144	4283	148	4283	148
5	4118	176	2482	144	4290	182	4135	148
6	4075	208	2482	144	4289	221	4135	148
7	4043	226	2482	144	4278	243	4135	148
8	4024	242	2482	145	4277	264	4135	328
9	4008	257	2482	146	4280	283	4135	356
10	3996	272	2482	146	4285	303	4135	356
11	3983	286	2482	147	4279	322	4135	417
12	3976	296	2482	148	4282	336	4135	424
13	3967	307	2482	149	4280	351	4135	439
14	3960	311	2482	150	4276	358	4135	461
15	3952	319	2482	150	4273	372	4135	461
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
30	3928	360	2482	152	4284	443	4135	543

Table 2.1: Number of patients who died at or before time t and the number of patients compatible with $\overline{a}(0)$ and $\overline{a}(5)$, before and after weighting; n_a and d_a refer to the denominator and numerator, respectively, of the naïve estimator (2.2) and the IPTW-estimator (2.3).

pooled logistic regression model for this purpose, following the strategy explained in (Vansteelandt et al., 2009). The probability of acquiring infection was associated with the presence of infection (other than NP) upon ICU admission and the following invasive treatments, received the day before infection (i.e., at time $t-1$): mechanical ventilation, parenteral feeding and feeding through naso- or oro- intestinal tube, tracheotomy, oral intubation, the presence/absence of tracheotomy intubation, surgery and a central vascular catheter; and antibiotic treatment received two days before infection (i.e., at time $t-2$). We did not adjust for antibiotic treatment the day before infection to exclude the possibility that this was already an effect of a latent infection. Aditionally there was an association with gender and study center, evidence for effect modification of the SAPS II score by type of admission, multiple trauma and surgery, and evidence for effect modification of the effect of antibiotic treatment and parenteral feeding by study center. Finally, we used regression splines for the time effect. Using the predicted values from the final model, we calculated the probability of each patient being compatible with the chosen infection

path up to time t, given baseline variables and time-dependent variables $\overline{L}_{t-1}$. The calculation of the resulting weights $W_{i\overline{a}t}$ is illustrated in Table 2.2 for three effectively observed patients who were discharged from the ICU at 5, 9 and 3 days after admission, respectively. In particular, the weight $W_{i\overline{a}t}$ is obtained by taking the reciprocal of the product of the probabilities π_{2is} over all time points $s = 3, ..., t$ (the remaining columns of Table 2.2 will be explained in Section 2.4.2). Note the sharp increase in the weights when patients become infected. It follows that an analysis of most infection paths, such as $\overline{a}(5)$, may be much influenced by extreme weights so that less reliable inferences are typically obtained under these infection paths. For instance, the weights of patients who are compatible with $\overline{a}(5)$ had a median and mean of 1.01 and 1.59, an interquartile range and standard deviation of 0.03 and 9.47 and 1% and 99% percentiles of 1 and 9.93 (min. 1, max. 387.4). In contrast, inference for the infection-free path $\overline{a}(0)$ tends not to suffer much from influential weights because most observed patient-days correspond to no infection. For instance, the weights of patients who are compatible with $\overline{a}(0)$ had a median and mean of 1.02 and 1.08, an interquartile range and standard deviation of 0.05 and 0.22 and 1% and 99% percentiles of 1 and 1.93 (min. 1, max. 9.64). This is useful because the most relevant intervention is one where infection would be prevented for all patients and thus our most substantive interest lies in $\overline{a}(0)$.

Using the calculated weights, we can now predict the number of patients who would have died within the ICU by a given time had infection been prevented for all. This is given by the numerator of (2.3) and is displayed in the 7th column of Table 2.1. The corresponding overall number of patients (i.e. the denominator of (2.3)) is shown in column 6. It approximates the overall sample size (i.e., 4288) because the idea behind inverse weighting is to infer how the data would have looked like had all patients received the same intervention (e.g. been uninfected). The last 2 columns of Table 2.1 show similar results for the infection path $\overline{a}(5)$. The IPTW-estimators (2.3) corresponding to $\overline{a}(0)$ and $\overline{a}(5)$ are again obtained as $d_{\overline{a}(0)}/n_{\overline{a}(0)}$ and $d_{\overline{a}(5)}/n_{\overline{a}(5)}$, respectively, and displayed in Figure 2.3(B) along with the observed CIF. Not surprisingly, the CIF corresponding to the infection-free path is now higher than before because it

i	t	A_{it}	π_{1it}	π_{2it}	π_{3it}	$W_{i\overline{a}t}$	$h_{t\overline{a}_t k}(L_{i0})$	$h_{t\overline{a}_t k}(L_{i0})W_{i\overline{a}t}$
1	3	0	0.98	0.99	0	1.01	0.98	0.99
1	4	1	0.02	0.01	0.36	100	0.03	3
1	5	1	1	1	0.29	100	0.03	3
2	3	0	0.98	0.99	0	1.01	0.98	0.99
2	4	0	0.98	0.99	0.36	1.02	0.97	0.99
2	5	1	0.02	0.02	0.29	50	0.05	2.5
2	6	1	1	1	0.23	50	0.05	2.5
2	7	1	1	1	0.18	50	0.05	2.5
2	8	1	1	1	0.15	50	0.05	2.5
2	9	1	1	1	0.12	50	0.05	2.5
3	3	0	0.98	0.99	0	1.01	0.98	1

Table 2.2: Estimated weights for 3 patients who are discharged on days 5, 9 and 3, respectively, corresponding to different infection histories; $\pi_{1it} = P(A_{it} = a_{it}|\epsilon_i(t) = 0, A_{i,t-1} = 0, L_{i0})$, $\pi_{2it} = P(A_{it} = a_{it}|\epsilon_i(t), A_{i,t-1} = 0, \overline{L}_{i,t-1})$, $\pi_{3it} = P(\epsilon_i(t) > 0|\epsilon_i(t-1) = 0, \overline{A}_{i,t-1}, L_{i0})$.

acknowledges that patients who avoid infection during their ICU stay are relatively more healthy. The IPTW-estimate of the CIF under infection path $\overline{a}(5)$ (Figure 2.3(B)) is now more realistic than the naïve one, although very imprecise (see Figure 2.4(B)). The fact that the IPTW-estimate of the CIF under $\overline{a}(0)$ suggests slightly higher ICU death rates than in the observed population, might be counterintuitive, although there is little that can be said about this comparison given the imprecision of these estimates. The more efficient inferences developed in the next section will shed further light on this.

2.4 Estimation under a semiparametric discrete-time marginal structural subdistribution hazard model

2.4.1 Model and estimation

The previous IPTW estimator is useful to infer the CIF of death that would be realized if all patients were uniformly exposed to the same infection path, but does not allow for direct comparison of CIFs corresponding to different chosen infection paths. It thus yields no effect estimates. Furthermore, in finite samples, the IPTW estimator is not guaranteed to yield a non-decreasing

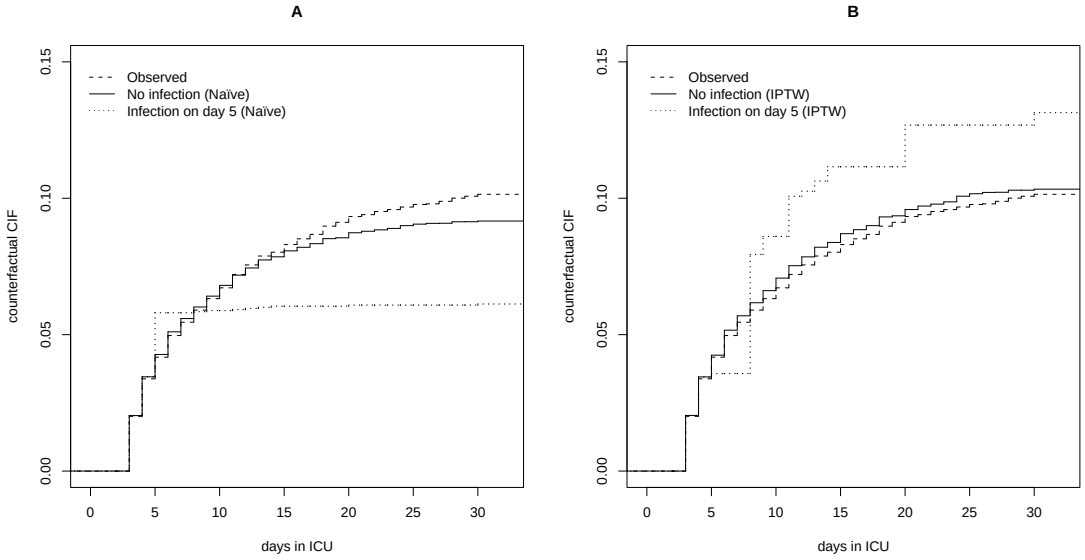

Figure 2.3: Naïve estimate (A) and IPTW-estimate (B) of the counterfactual CIF for ICU death under infection paths $\overline{a}(0)$ and $\overline{a}(5)$, together with the observed CIF.

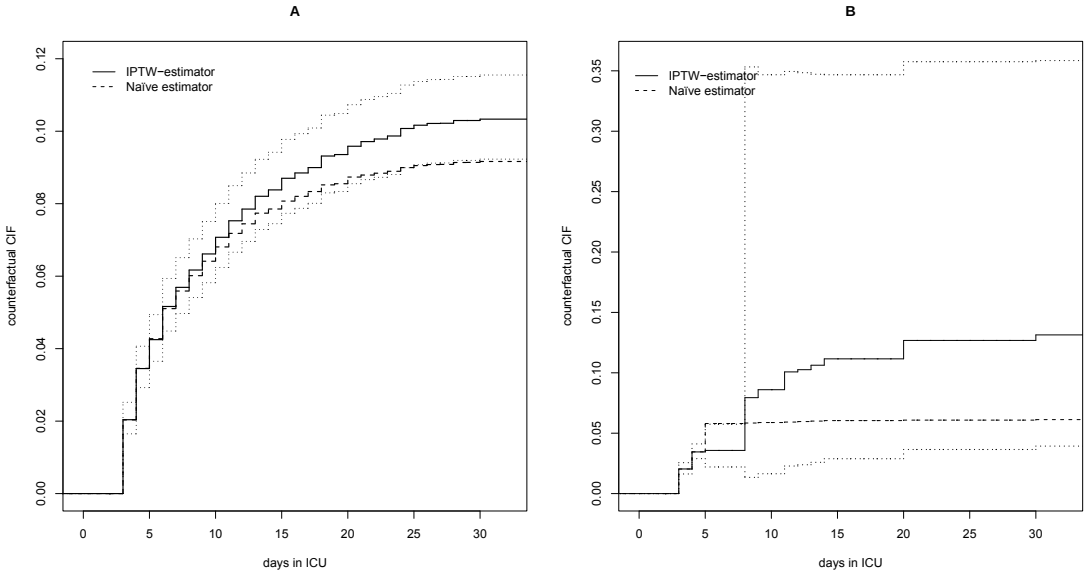

Figure 2.4: IPTW-estimate (with 95% pointwise confidence intervals) and naïve estimate of the counterfactual CIF for ICU death under infection paths $\overline{a}(0)$ (A) and $\overline{a}(5)$ (B).

counterfactual CIF with time and can be very inefficient (see e.g. Figure 2.4(B)).

In view of this, we will develop models and accompanying inference for the counterfactual subdistribution hazard corresponding to event type k, which we define conditional on baseline covariates L_0 as:

$$\lambda_{\overline{a}k}(t|L_0) = P(T_{\overline{a}} = t, \epsilon_{\overline{a}} = k | T_{\overline{a}} \geq t \cup (T_{\overline{a}} < t \cap \epsilon_{\overline{a}} \neq k), L_0\}.$$

For $k = 1$, this represents the risk for a patient to die within the ICU on day t under infection path $\overline{a}$, given that he/she did not die in the ICU before that day. Note that the risk set at each time t is composed of those who did not experience the event of interest (i.e., who did not die) by time t, thus including those who experienced a competing event (i.e., who got discharged) by time t. We can now define a discrete-time marginal structural subdistribution hazard model for event type k to be a model for the counterfactual discrete-time hazard, possibly conditional on baseline covariates L_0.

For example, with $\delta_t \equiv \sum_{s=1}^{t-1} a_s$ denoting the number of days a patient would have been infected prior to time t under infection path $\overline{a}$, one might postulate that

$$\text{logit}\left\{\lambda_{\overline{a}k}(t|L_0)\right\} = \omega(t) + \beta_1 a_t + \beta_2 \delta_t + \beta_3 L_0, \tag{2.7}$$

where $\omega(t)$ represents an unknown function of time which will later be modelled using regression splines and where β_1, β_2 and β_3 are unknown finite-dimensional parameters. When $k = 1$ and, as usual, the counterfactual hazard of death is relatively small at any given time, then $\exp\left(\beta_1 a_t + \beta_2 \delta_t\right)$ approximates the causal subdistribution rate ratio of ICU death at time t due to infection path $\overline{a}$ (as compared to the no infection path).

If, for each patient in the sample, we had the ability to measure $\epsilon_{\overline{a}}$ for each potential infection

path $\overline{a}$, then we could fit model (2.7) by solving the following estimating equation

$$0 = \sum_{i=1}^{n} \sum_{t=1}^{E} \sum_{\overline{a}_t} U_{it\overline{a}_t k} \left(h_{t\overline{a}_t k}; \beta \right), \tag{2.8}$$

where $\beta = (\beta_1 \ \beta_2 \ \beta_3)'$ and

$$\begin{aligned} U_{it\overline{a}_t k} \left(h_{t\overline{a}_t k}; \beta \right) \equiv & I(\epsilon_{i\overline{a}}(t-1) \neq k) h_{t\overline{a}_t k}(L_{i0}) \Gamma \\ & \times \left[I(\epsilon_{\overline{a}}(t) = k) - \text{expit} \left\{ \omega(t) + \beta_1 a_t + \beta_2 \delta_t + \beta_3 L_{i0} \right\} \right]. \end{aligned}$$

Here, $h_{t\overline{a}_t k}(L_{i0})$ is an arbitrary scalar function of $(t, \overline{a}_t, L_{i0})$, e.g. the constant 1, and Γ is the vector of predictors in the discrete-time marginal structural subdistribution hazard model, e.g. $(1 \ t \ a_t \ \delta_t \ L_{i0})'$ if $\omega(t)$ is chosen to be a linear function of time (i.e., $\omega(t) = \beta_0 + \beta_4 t$ with β_0 and β_4 unknown). The above estimating equation is not feasible because $\{\epsilon_{\overline{a}}(t), t = 1, ..., E\}$ is not measured for each patient under each potential infection path. Thinking along the lines explained in Section 2.3, we will therefore select for each considered infection path, those patients whose observed data are compatible with that path and use inverse probability weighting to account for the selective nature of those subgroups. This is realized by solving the following equations,

$$0 = \sum_{i=1}^{n} \sum_{t=1}^{E} \sum_{\overline{a}_t} U_{it\overline{a}_t k} \left(h_{t\overline{a}_t k}; \beta \right) W_{i\overline{a}t} \tag{2.9}$$

which only involve observed data.

The choice of $h_{t\overline{a}_t k}(L_{i0})$ can be very important in the analysis as it links directly to the efficiency of the estimator $\hat{\beta}$ for β obtained by solving this equation. In the Appendix, we show that for a single infection path $\overline{a}$, a locally efficient estimator within the above class is obtained by

choosing

$$h_{t\overline{a}_t k}(L_0) = \left[\sum_{d=1}^{t} P(\epsilon(d) \neq 0 | \overline{\epsilon(d-1)}, \epsilon(d-1) \neq 1, \overline{A}_{d-1} = \overline{a}_{d-1}, L_0) \right. \\ \left. \times \frac{\prod_{s=1}^{d-1} P(\epsilon(s) = 0 | \epsilon(s-1) = 0, \overline{A}_{s-1} = \overline{a}_{s-1}, L_0)}{\prod_{s=1}^{d-1} P(A_s = a_s | \epsilon(s) = 0, \overline{A}_{s-1} = \overline{a}_{s-1}, L_0)} \right]^{-1}$$

where, for each t, we define

$$P(\epsilon(d) \neq 0 | \overline{\epsilon(d-1)}, \epsilon(d-1) \neq 1, \overline{A}_{d-1} = \overline{a}_{d-1}, L_0) = 1$$

if $d = t + 1$.

Here, the efficiency is local in the sense that it is only attained when infection and discharge/death are independent of the considered time-varying confounders, conditional on the considered baseline confounders. The index function $h_{t\overline{a}_t k}(L_{i0})$ can be interpreted as the reciprocal of a prediction of the considered subject's inverse weight $W_{i\overline{a}t}$ in equation (2.9) on the basis of time, the considered infection path and baseline confounders. As such, using this index function has the effect of stabilizing the weights in the sense of making patients who get assigned small probabilities $P(A_t = a_t | \epsilon(t) = 0, \overline{A}_{t-1}, \overline{L}_{t-1})$, less influential in the analysis. In line with others (Hernán et al., 2000), we will refer to $W_{i\overline{a}t}$ as an unstable weight and to $h_{t\overline{a}_t k}(L_{i0}) W_{i\overline{a}t}$ as a stable weight.

In principle, the third summation in (2.8) runs over all possible infection paths that are considered to be relevant to the investigator. In practice, depending on the chosen discrete-time marginal structural subdistribution hazard model, the estimating functions $U_{t\overline{a}_t k}\left(h_{t\overline{a}_t k}; \beta\right)$ for a given subject at a given time t may be identical with probability 1 between a number of infection paths. For instance, suppose that β_2 were set to zero in model (2.7). Then all infection paths with $a_t = 1$ (or $a_t = 0$) yield the same estimating function value $U_{t\overline{a}_t k}\left(h_{t\overline{a}_t k}; \beta\right)$ at time

t, regardless of the infection history $\overline{a}_{t-1}$ or future $\underline{a}_{t+1} \equiv (a_{t+1}, \ldots, a_E)$. In view of this, we recommend to sum at each time t only over those infection paths in the estimating equation (2.9) that yield unique estimating function values (with probability 1) at that time. This is inspired by the fact that the efficient score for the parameters indexing a conditional mean model counts perfectly correlated observations only once (Chamberlain, 1987). The resulting analysis continues to yield a consistent (and likely more efficient) estimator because it follows from the Appendix that each term in the summation (2.9) has mean zero when evaluated at the truth.

A conservative estimate for the asymptotic variance of the estimator $\hat{\beta}$ which solves equation (2.9) can be obtained via the sandwich estimator

$$\frac{1}{n} E^{-1} \left(\frac{\partial U_i(\beta)}{\partial \beta} \right) Var\left\{ U_i(\beta) \right\} E^{-1} \left(\frac{\partial U_i(\beta)}{\partial \beta} \right)',$$

where we define

$$U_i(\beta) = \sum_{t=1}^{E} \sum_{\overline{a}_t} U_{it\overline{a}_t k}\left(h_{t\overline{a}_t k}; \beta\right) W_{i\overline{a}t}$$

and where population expectations and variances can be replaced with sample analogs and β with $\hat{\beta}$. The counterfactual CIF corresponding to infection path $\overline{a}$ among patients with baseline covariate $L_0 = l_0$ can be estimated under the discrete-time marginal structural subdistribution hazard model (2.7) as

$$\hat{F}_{\overline{a}k}(t) = 1 - \prod_{s \leq t} \left\{ 1 - \lambda_{\overline{a}k}(s|l_0) \right\}.$$

This counterfactual CIF is always well-defined because the marginal structural subdistribution hazard model (2.7) only involves baseline covariates and no time-varying covariates. Using the Delta method, a corresponding conservative, asymptotic $100(1-\alpha)\%$ confidence interval for given t is obtained as

$$1 - \exp\left[\log\{1 - \hat{F}_{\overline{a}k}(t)\} \pm z_{\alpha/2} \sigma_{\overline{a}k}(t) \right]$$

where

$$\sigma^2_{\overline{a}k}(t) = \left(\sum_{s \leq t} \frac{\lambda'_{\overline{a}k}(s|l_0)}{1 - \lambda_{\overline{a}k}(s|l_0)} \right) Var(\hat{\beta}) \left(\sum_{s \leq t} \frac{\lambda'_{\overline{a}k}(s|l_0)}{1 - \lambda_{\overline{a}k}(s|l_0)} \right)'$$

and where $\lambda'_{\overline{a}k}(s|l_0)$ is the first order derivative of $\lambda_{\overline{a}k}(s|l_0)$ w.r.t. β.

2.4.2 Data analysis

The stable weights $h_{t\overline{a}_t k}(L_{i0})W_{i\overline{a}t}$ were obtained by estimating $h_{t\overline{a}_t k}(L_{i0})$ based on two pooled logistic regression models for discharge/death and infection in function of time. In our analysis, the model for discharge/death contained the infection history and regression splines for the time effect, but (for computational simplicity) no baseline covariates. Table 2.2 illustrates the calculation of the stable weights for three effectively observed patients. Note indeed that the resulting weights in the last column of Table 2.2 are more stable than the unstable weights $W_{i\overline{a}t}$ that were previously considered. It can be shown that the unstable weights stay constant after discharge and that the stable weights for a given infection path $\overline{a}$ remain constant at those times t where $a_t = 1$. In general, the stable weights continue to change with time after discharge because they involve the probability of getting discharged or infected at each time.

	Naïve		
	Estimate	SE	OR
β_1	-3.27^a	0.21	0.038
β_2	0.12^a	0.022	1.13

Table 2.3: The naïve parameter estimates for model (2.7) including (the most important) baseline covariates;^{a} significant at the 5% significance level.

Parameter estimates of the infection effect were obtained by solving the weighted estimating equations (2.9). This required extending the dataset by including all compatible infection paths, as illustrated in Figure 2.2. We estimate the time-dependent intercept $\omega(t)$ in model (2.7) using natural cubic splines with five knots. The analysis was adjusted for baseline covariates (ICU center, SAPSII score on admission, infection upon admission, oral intubation on admission and

	IPTW ($W_{i\bar{a}t}$)			IPTW ($h_{t\bar{a}_t k}(L_{i0})W_{i\bar{a}t}$)		
	Estimate	SE	OR	Estimate	SE	OR
β_1	-0.33	0.35	0.72	-0.49	0.28	0.61
β_2	0.095^a	0.035	1.10	0.095^a	0.024	1.10
β_2	0.087^a	0.032	1.09	0.059^a	0.021	1.06

Table 2.4: IPTW-estimates for model (2.7) including (the most important) baseline covariates using unstable and stable weights; ^{a} significant at the 5% significance level.

type of admission), which were also used in the infection models in the calculation of the stable weights. To check the stability of the weights for a given patient at a given time t we calculated that patient's overall weight at that time as the sum of the weights over all compatible infection paths. The resulting stable weights had a median and mean of 2.37 and 2.74, an interquartile range and standard deviation of 2.57 and 1.97 and 1% and 99% percentiles of 0.41 and 7.35 (min. 0.16, max. 53.4). In contrast, the unstable weights had a median and mean of 13.08 and 15.71, an interquartile range and standard deviation of 15.91 and 25.19 and 1% and 99% percentiles of 1 and 90.36 (min. 1, max. 529.81).

The results for model (2.7) are given in the first two rows of Table 2.3 and 2.4. Note the large difference between the estimates obtained with and without inverse weighting. This underscores the importance of confounding adjustment. The IPTW-estimates for β_1 suggest a drop in the hazard of death during the first 4-5 days after acquiring infection (which could possibly be explained by extra care after infection or delayed decision of treatment interruption), but the evidence for this is weak. We hence fitted a reduced model by setting $\beta_1 = 0$ and found a 6.1% (95%CI 1.8 - 10.0) increase in the hazard of ICU death per additional day since acquiring infection (see the third row of Table 2.3). This reflects the increased risk of death which is partly attributable to a direct effect of infection on death, and partly to the increase in length of stay after infection, which implies a greater chance of 'observing' death. Figure 2.5 shows the estimated counterfactual CIFs obtained using the naïve and IPTW-estimates from the final models (but now no longer conditioning on baseline covariates). They give similar results as obtained in Section 2.3

using the nonparametric estimators, but a drastic increase in efficiency for the infection paths. In particular, the stable IPTW-estimator suggests a roughly 3% increase in ICU mortality due to infection on day 5 as compared to no infection. Figure 2.6 allows for evaluating how much the CIF of ICU death would change if all infections could be avoided. It suggests a population attributable fraction of 30-day ICU mortality related to infection of 6.4%, indicating that 6.4% of the observed ICU deaths at 30 days could have been avoided by preventing infection.

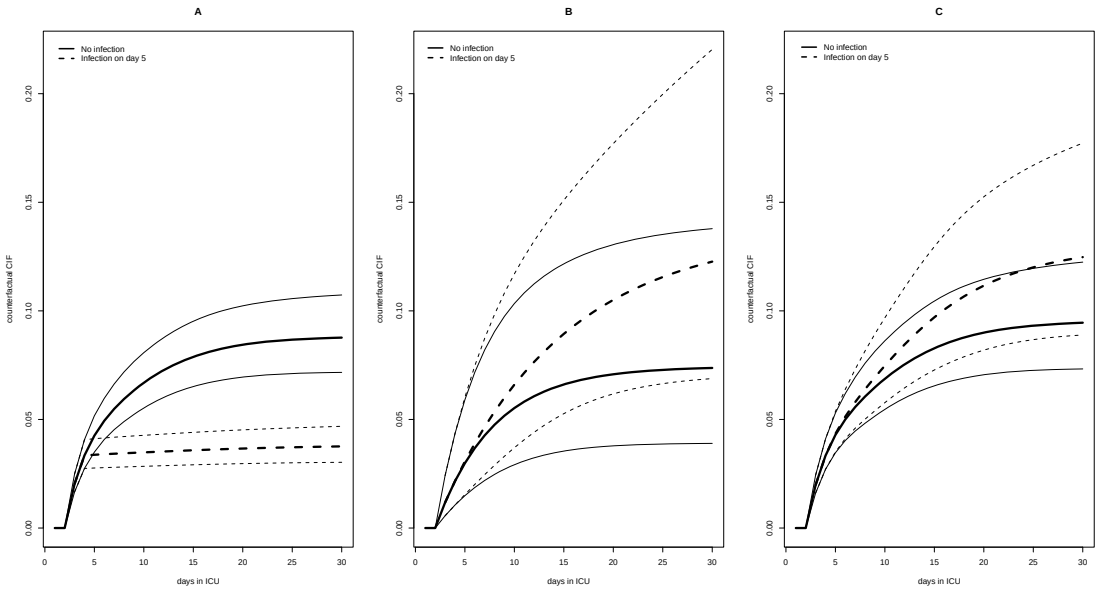

Figure 2.5: Counterfactual CIF of ICU death for infection paths $\overline{a}(0)$ and $\overline{a}(5)$, as based on the naïve estimator (A), the unstable IPTW-estimator (B) and the stable IPTW-estimator (C).

2.5 Discussion

We have developed a general framework for estimating the causal effect of a time-varying exposure on a survival outcome in the presence of competing risks. The validity of the approach relies on the following four assumptions. First, it assumes correctly specified models for the probability of infection at each time in function of the history of measured time-varying confounders. Second, it assumes that there are no patients who, on the basis of their time-varying

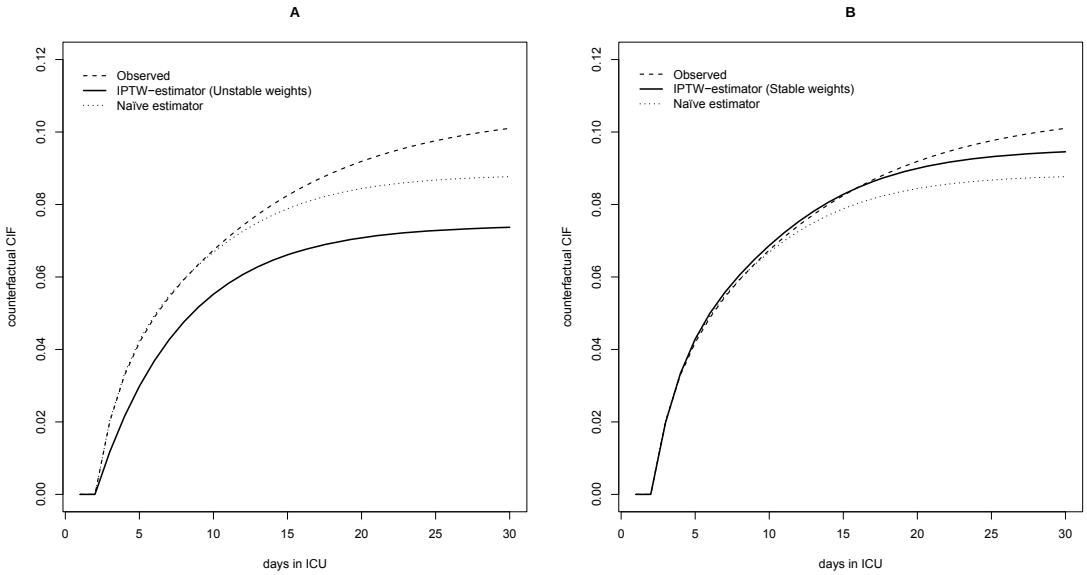

Figure 2.6: Counterfactual CIF of ICU death for infection paths $\overline{a}(0)$, as based on the naïve estimator, the unstable IPTW-estimator (A) and the stable IPTW-estimator (B) together with the observed CIF of death.

confounder history, are precluded from avoiding or acquiring infection. While we believe the second assumption was realistic in our data analysis, in general, there may be patients who have a small daily risk of infection on the basis of their time-varying confounder history. This can lead to patients having large weights in the analysis who may thus become influential.

Both previous assumptions are (at least to some degree) testable on the basis of the observed data. The next two assumptions are inherently untestable, so that their plausibility must be judged against subject-matter knowledge. First, the proposed approach assumes that at each time, all time-varying confounders for the association between infection and ICU mortality have been measured. While this assumption might be approximately justified for our analysis, it is likely not entirely met because we could only adjust for the applied invasive treatments, which should be regarded as a surrogates for the actual disease severity. In next chapter, we applied the proposed methods to the Outcomerea (French multicenter ICU database) database which

carry very extensive information about the patients' severity of illness. Second, the approach additionally assumes that confounders measured at time $t-1$, i.e. L_{t-1}, are a cause, rather than a consequence of infection at time t. This assumption may not be satisfied when infection is in incubation several days prior to its detection.

An important advantage of our approach is that it separates the adjustment for time-varying confounders from the final model for the counterfactual CIF. This implies that this approach can technically handle high-dimensional confounders without imposing a high computational burden and even when confounders are themselves influenced by the infection. In addition, because the model involves no time-varying covariates, the corresponding CIF is well defined and can be displayed for the overall patient population, even with adjustment for high-dimensional confounders, without the need to restrict the CIF to patient subgroups. A further advantage is that, because infection paths are well defined upon ICU admission, our approach does not require imputing or stopping the covariate process (including infection status) for patients who experience the competing event (Beyersmann and Schumacher, 2008; Latouche et al., 2005). A final advantage is that it yields fairly easily interpretable results. For instance, the results can be expressed in terms of population attributable fractions comparing the observed population with how that same population would have looked like if infection were prevented for all patients. However, caution must be exercised because the results state the impact of hospital-acquired infections on ICU mortality rather than overall mortality. This implies that the magnitude of the infection effect is partly influenced by e.g. the decision to keep infected patients longer in the ICU (which implies a greater chance of observing ICU death) or to send terminally ill patients home. Care must also be exercised in noting that late infections are forced to have a small effect under our model because most patients will have left the ICU by that time and thus not be affected. This explains the discrepancy with the results of (Vansteelandt et al., 2009), who estimate the effect of acquiring infections while still being present in the ICU.

We did not discuss how to deal with censored event times as this was not an issue in our data analysis. In the presence of the usual right-censoring, our method can be extended using additional inverse probability of censoring weights along the lines of Fine and Gray (Robins and Rotnitzky, 1992; Fine and Gray, 1999). Finally, we have chosen a development which adopts the principles underlying marginal structural models (Robins et al., 2000). Alternatively, we could have considered a development based on structural nested failure time models, as in (Schulgen and Schumacher, 1996; Keiding et al., 1999). We have chosen not to do so because the survival time of critically ill patients can be greatly influenced by the decision of doctors and family to keep patients artificially alive, so that 30-day ICU mortality tends to be a more relevant endpoint to physicians. In addition, it is so far unclear how to account for censoring due to ICU discharge in this approach, without running into the aforementioned difficulties caused by inverse probability of censoring weighting.

In conclusion, we hope that the proposed methods will shed new light on addressing issues of time-dependent confounding in survival analyses with competing risks and, more specifically, will help improve our understanding of the role of nosocomial infections in ICU mortality.

Appendix

Derivation of expression (2.5)

We drop the subject index i for notational simplicity. To derive expression (2.5), note that $\hat{F}_{\overline{a}k}(t)$ is such that $\sum_{i=1}^{n} U_{it\overline{a}_t k} = 0$, with $U_{it\overline{a}_t k}$ defined in (2.6). It then follows using standard arguments for M-estimators that, under weak regularity conditions and for given t, the asymptotic variance of $\hat{G}_{\overline{a}k}(t) \equiv \text{logit}\left\{\hat{F}_{\overline{a}k}(t)\right\}$ equals

$$\text{Var}(\hat{G}_{\overline{a}k}(t)) = \frac{1}{n} E^{-2}\left(\frac{\partial U_{t\overline{a}_t k}}{\partial G_{\overline{a}k}(t)}\right) \text{Var}(U_{t\overline{a}_t k}),$$

where

$$E\left(\frac{\partial U_{t\overline{a}_t k}}{\partial G_{\overline{a}k}(t)}\right) = F_{\overline{a}k}(t)\{1 - F_{\overline{a}k}(t)\}.$$

The result (2.5) is now immediate. It yields a conservative $(1-\alpha)100\%$ confidence interval by the fact that it does not acknowledge estimation of the unknown population probabilities appearing in the inverse weights (van der Laan and Robins, 2003).

Unbiasedness of estimating equation (2.9)

For simplicity of notation, we write $U_{t\overline{a}_t}(\beta) \equiv U_{t\overline{a}_t k}(h_{t\overline{a}_t k}; \beta)$ for a given choice of $h_{t\overline{a}_t k}(L_{i0})$. Using the law of iterated expectations, the population expectation of (2.9), evaluated at β, equals

$$\begin{aligned}
&E\left[\sum_{t=1}^{E}\sum_{\overline{a}_t} E\left(U_{t\overline{a}_t}(\beta)\prod_{s=1}^{t}\frac{I(A_s = a_s)I(\epsilon(s)=0) + I(\epsilon(s)\neq 0)}{P(A_s = a_s|\epsilon(s), \overline{A}_{s-1}, \overline{L}_{s-1})} | \overline{A}_{t-1}, \overline{L}_{t-1}, \overline{\epsilon_{\overline{a}}(t)}\right)\right] \\
&= E\left[\sum_{t=1}^{E}\sum_{\overline{a}_t} U_{t\overline{a}_t}(\beta)\prod_{s=1}^{t-1}\frac{I(A_s = a_s)I(\epsilon(s)=0) + I(\epsilon(s)\neq 0)}{P(A_s = a_s|\epsilon(s), \overline{A}_{s-1}, \overline{L}_{s-1})}\right. \\
&\left.\times E\left(\frac{I(A_t = a_t)I(\epsilon(t)=0) + I(\epsilon(t)\neq 0)}{P(A_t = a_t|\epsilon(t), \overline{A}_{t-1}, \overline{L}_{t-1})} \mid \overline{A}_{t-1}, \overline{L}_{t-1}, \overline{\epsilon_{\overline{a}}(t)}\right)\right]. \qquad (2.10)
\end{aligned}$$

Here, the expectation in the last line can be rewritten as

$$\frac{P(A_t = a_t|\overline{\epsilon_{\overline{a}}(t)}, \overline{A}_{t-1}, \overline{L}_{t-1})I(\epsilon(t) = 0)}{P(A_t = a_t|\epsilon(t), \overline{A}_{t-1}, \overline{L}_{t-1})} + \frac{I(\epsilon(t) \neq 0)}{P(A_t = a_t|\epsilon(t), \overline{A}_{t-1}, \overline{L}_{t-1})}. \tag{2.11}$$

The first term in (2.11) equals $I(\epsilon(t) = 0)$ under the no unmeasured confounders assumption. The second term in (2.11) equals $I(\epsilon(t) \neq 0)$ by the fact that we defined $P(A_t = a_t|\overline{\epsilon_{\overline{a}}(t)}, \overline{A}_{t-1}, \overline{L}_{t-1}) = P(A_t = a_t|\epsilon(t), \overline{A}_{t-1}, \overline{L}_{t-1})$ to equal 1 for subjects with $\epsilon(t) \neq 0$. Equation (2.10) thus simplifies to

$$E\left[\sum_{t=1}^{T}\sum_{\overline{a}_t} U_{t\overline{a}_t}(\beta)\prod_{s=1}^{t-1}\frac{I(A_s = a_s)I(\epsilon(s) = 0) + I(\epsilon(s) \neq 0)}{P(A_s = a_s|\epsilon(s), \overline{A}_{s-1}, \overline{L}_{s-1})}\right]$$

Repeating these arguments another $t-1$ times, we obtain $E\left\{\sum_{t=1}^{T}\sum_{\overline{a}_t} U_{t\overline{a}_t}(\beta)\right\}$ which has mean zero by the fact that $U_{t\overline{a}_t}(\beta)$ is an unbiased estimating function for each t and $\overline{a}_t$. A similar proof can be developed, starting from estimating function (2.6) to show that (2.3) is a consistent estimator of the CIF under infection path $\overline{a}$ at time t.

Stable weights

For simplicity of notation and without loss of generality, we let $k = 1$ and drop this index in all further expressions. To derive stabilized weights, we focus on 1 infection path $\overline{a}$ and calculate the efficient index function $h_{t\overline{a}_t}(L_0)$ in the estimating function

$$\sum_{t=1}^{T} h_{t\overline{a}_t}(L_0)V_{t\overline{a}_t}W_{t\overline{a}_t}, \tag{2.12}$$

where we define, for notational convenience,

$$\begin{aligned} V_{t\overline{a}_t} &\equiv \left\{I(\epsilon_{\overline{a}}(t) = 1) - \text{expit}(\beta' X)\right\} I(\epsilon_{\overline{a}}(t-1) \neq 1) \\ W_{t\overline{a}_t} &\equiv \prod_{s=1}^{t}\frac{I(A_s = a_s)I(\epsilon(s) = 0) + I(\epsilon(s) \neq 0)}{P(A_s = a_s|\epsilon(s), \overline{A}_{s-1}, \overline{L}_{s-1})}, \end{aligned}$$

where X is, for instance, equal to $(1\ t\ a_t\ \delta_t\ L_0)'$. It follows from standard semiparametric theory (Chamberlain, 1987) that the efficient index function (in the sense of leading to an estimator of β with minimal variance among all estimators obtained by solving an estimating equation based on estimating functions of the form (2.12) with $P(A_s = a_s|\epsilon(s), \overline{A}_{s-1}, \overline{L}_{s-1})$ known for $s = 1, \ldots E$) is given by

$$h_{t\overline{a}_t}(L_0) = E\left(\frac{\partial}{\partial\beta}V_{t\overline{a}_t}W_{t\overline{a}_t}|L_0\right)\mathrm{Var}^{-1}\left(V_{t\overline{a}_t}W_{t\overline{a}_t}|L_0\right). \tag{2.13}$$

Define $\epsilon_{t\overline{a}} \equiv (I(\epsilon_{\overline{a}}(1) \neq 1), ..., (\epsilon_{\overline{a}}(t) \neq 1))$. Using similar arguments as in the previous appendix, one can show that $E\left(\frac{\partial}{\partial\beta}V_{t\overline{a}_t}W_{t\overline{a}_t}|L_0\right)$ equals $X\mathrm{expit}(\beta'X)\{1-\mathrm{expit}(\beta'X)\}$ $\times P(\epsilon_{\overline{a}}(t-1) \neq 1|L_0)$. Using the law of iterated variances, the variance in (2.13) can be calculated as

$$E\left\{\mathrm{Var}\left(V_{t\overline{a}_t}W_{t\overline{a}_t}|\epsilon_{t\overline{a}}, L_0\right)|L_0\right\} + \mathrm{Var}\left\{E\left(V_{t\overline{a}_t}W_{t\overline{a}_t}|\epsilon_{t\overline{a}}, L_0\right)|L_0\right\}. \tag{2.14}$$

The second term in (2.14) equals

$$\mathrm{expit}(\beta'X)\{1-\mathrm{expit}(\beta'X)\}P(\epsilon_{\overline{a}}(t-1) \neq 1|L_0)E\left(W_{t\overline{a}_t}|\epsilon_{t\overline{a}}, L_0\right)^2.$$

The first term in (2.14) equals

$$E\left\{V^2_{t\overline{a}_t}\mathrm{Var}\left(W_{t\overline{a}_t}|\epsilon_{t\overline{a}}, L_0\right)|L_0\right\} = E\left[V^2_{t\overline{a}_t}\left\{E\left(W^2_{t\overline{a}_t}|\epsilon_{t\overline{a}}, L_0\right) - E\left(W_{t\overline{a}_t}|\epsilon_{t\overline{a}}, L_0\right)^2\right\}|L_0\right].$$

We first calculate $E\left(W^2_{t\overline{a}_t}|\overline{A}_{t-1}, \overline{L}_{t-1}, \epsilon_{t\overline{a}}\right)$ as

$$E\left[W^2_{t-1,\overline{a}_{t-1}}\frac{I(A_t = a_t)I\left(\epsilon(t) = 0\right) + I\left(\epsilon(t) \neq 0\right)}{P(A_t = a_t|\epsilon(t), \overline{A}_{t-1}, \overline{L}_{t-1})^2}|\overline{A}_{t-1}, \overline{L}_{t-1}, \epsilon_{t\overline{a}}\right]$$
$$= E\left\{W^2_{t-1,\overline{a}_{t-1}}\left[\frac{I\left(\epsilon(t) = 0\right)}{P(A_t = a_t|\epsilon(t) = 0, \overline{A}_{t-1} = \overline{a}_{t-1}, \overline{L}_0)} + I\left(\epsilon(t) \neq 0\right)\right]|\overline{A}_{t-1}, \overline{L}_{t-1}, \epsilon_{t\overline{a}}\right\},$$

where we use the no unmeasured confounders assumption along with the simplifying assumption that $A_t \perp\!\!\!\perp \overline{L}_{t-1}|\epsilon(t) = 0, \overline{A}_{t-1}, L_0$ (note that violation of this assumption may affect the efficiency, but not the asymptotic unbiasedness of the estimator because each choice of $h_{t\overline{a}_t}(L_0)$ yields a consistent estimator). Starting from here, we next calculate $E\left(W^2_{t\overline{a}_t}|\overline{A}_{t-1}, \overline{L}_{t-2}, \epsilon_{t\overline{a}}\right)$ as

$$
\begin{aligned}
&E\left(W^2_{t-1,\overline{a}_{t-1}}\left[\frac{I\left(\epsilon(t)=0\right)}{P(A_t=a_t|\epsilon(t)=0,\overline{A}_{t-1}=\overline{a}_{t-1},\overline{L}_0)^2}+I\left(\epsilon(t)\neq 0\right)\right]|\overline{A}_{t-1},\overline{L}_{t-1},\epsilon_{t\overline{a}}\right)\\
&=E\left\{W^2_{t-1,\overline{a}_{t-1}}\left(I\left(\epsilon(t-1)=0\right)\left\{\frac{P(\epsilon(t)=0|\epsilon(t-1)=0,\overline{A}_{t-1}=\overline{a}_{t-1},L_0)}{P(A_t=a_t|\epsilon(t)=0,\overline{A}_{t-1}=\overline{a}_{t-1},\overline{L}_0)}\right.\right.\right.\\
&\left.\left.\left.+P(\epsilon(t)\neq 0|\epsilon(t-1)=0,\overline{A}_{t-1}=\overline{a}_{t-1},L_0)\right\}+I\left(\epsilon(t-1)\neq 0\right)\right)|\overline{A}_{t-1},\overline{L}_{t-1},\epsilon_{t\overline{a}}\right\}
\end{aligned}
$$

where we now make the simplifying assumption that $P(\epsilon(t)|\epsilon(t-1)=0,\overline{A}_{t-1}=\overline{a}_{t-1},\overline{L}_{t-1},\epsilon_{t\overline{a}}) = P(\epsilon(t)|\epsilon(t-1)=0,\overline{A}_{t-1}=\overline{a}_{t-1},L_0)$ (note again that violation of this assumption may affect the efficiency, but not the asymptotic unbiasedness of the estimator). Starting from here, we then calculate $E\left(W^2_{t\overline{a}_t}|\overline{A}_{t-2}, \overline{L}_{t-2}, \epsilon_{t\overline{a}}\right)$ as

$$
\begin{aligned}
&E\left\{W^2_{t-2,\overline{a}_{t-2}}\left[I\left(\epsilon(t-1)\neq 0\right)+\frac{I(A_{t-1}=a_{t-1})I\left(\epsilon(t-1)=0\right)}{P(A_{t-1}=a_{t-1}|\epsilon(t-1)=0,\overline{A}_{t-2}=\overline{a}_{t-2},\overline{L}_{t-2})^2}\right.\right.\\
&\left.\left.\times\left\{\frac{P(\epsilon(t)=0|\epsilon(t-1)=0,\overline{A}_{t-1}=\overline{a}_{t-1},L_0)}{P(A_t=a_t|\epsilon(t)=0,\overline{A}_{t-1}=\overline{a}_{t-1},\overline{L}_0)}+P(\epsilon(t)\neq 0|\epsilon(t-1)=0,\overline{A}_{t-1}=\overline{a}_{t-1},L_0)\right\}\right]\right.\\
&\left.|\overline{A}_{t-2},\overline{L}_{t-2},\epsilon_{t\overline{a}}\right\}\\
&=E\left\{W^2_{t-2,\overline{a}_{t-2}}\left[I\left(\epsilon(t-1)\neq 0\right)+\frac{I\left(\epsilon(t-1)=0\right)}{P(A_{t-1}=a_{t-1}|\epsilon(t-1)=0,\overline{A}_{t-2}=\overline{a}_{t-2},\overline{L}_{t-2})}\right.\right.\\
&\left.\left.\times\left\{\frac{P(\epsilon(t)=0|\epsilon(t-1)=0,\overline{A}_{t-1}=\overline{a}_{t-1},L_0)}{P(A_t=a_t|\epsilon(t)=0,\overline{A}_{t-1}=\overline{a}_{t-1},\overline{L}_0)}+P(\epsilon(t)\neq 0|\epsilon(t-1)=0,\overline{A}_{t-1}=\overline{a}_{t-1},L_0)\right\}\right]\right.\\
&\left.|\overline{A}_{t-2},\overline{L}_{t-2},\epsilon_{t\overline{a}}\right\}
\end{aligned}
$$

where we use the simplifying assumption that $A_{t-1} \perp\!\!\!\perp \overline{L}_{t-2} | \epsilon(t-1) = 0, \overline{A}_{t-2}, L_0$. Continuing along these lines, we obtain that

$$E\left(W^2_{t\overline{a}_t} | \epsilon_{t\overline{a}_t}, L_0\right) = \sum_{d=1}^{t+1} P(\epsilon(d) \neq 0 | \epsilon(d-1) = 0, \overline{A}_{d-1} = \overline{a}_{d-1}, L_0) \times \frac{\prod_{s=1}^{d-1} P(\epsilon(s) = 0 | \epsilon(s-1) = 0, \overline{A}_{s-1} = \overline{a}_{s-1}, L_0)}{\prod_{s=1}^{d-1} P(A_s = a_s | \epsilon(s) = 0, \overline{A}_{s-1} = \overline{a}_{s-1}, L_0)},$$

where, for each t, we define $P(\epsilon(t+1) \neq 0 | \epsilon(t) = 0, \overline{A}_t = \overline{a}_t, L_0) = 1$. A similar calculation shows that $E\left(W_{t\overline{a}_t} | \epsilon_{t\overline{a}}, L_0\right) = 1$. It thus follows that (2.14) equals

$$\mathrm{Var}\left(V_{t\overline{a}_t} | L_0\right) \sum_{d=1}^{t+1} P(\epsilon(d) \neq 0 | \overline{\epsilon(d-1)}, \epsilon(d-1) \neq 1, \overline{A}_{d-1} = \overline{a}_{d-1}, L_0) \times \frac{\prod_{s=1}^{d-1} P(\epsilon(s) = 0 | \epsilon(s-1) = 0, \overline{A}_{s-1} = \overline{a}_{s-1}, L_0)}{\prod_{s=1}^{d-1} P(A_s = a_s | \epsilon(s) = 0, \overline{A}_{s-1} = \overline{a}_{s-1}, L_0)}$$

and that $h_{t,\overline{a}_t}(L_0)$ in (2.13) equals

$$X \left[\sum_{d=1}^{t+1} P(\epsilon(d) \neq 0 | \overline{\epsilon(d-1)}, \epsilon(d-1) \neq 1, \overline{A}_{d-1} = \overline{a}_{d-1}, L_0) \times \frac{\prod_{s=1}^{d-1} P(\epsilon(s) = 0 | \epsilon(s-1) = 0, \overline{A}_{s-1} = \overline{a}_{s-1}, L_0)}{\prod_{s=1}^{d-1} P(A_s = a_s | \epsilon(s) = 0, \overline{A}_{s-1} = \overline{a}_{s-1}, L_0)} \right]^{-1}.$$

Derivation of variance of expression 2.1

The attributable fraction is defined as:

$$\frac{F_1(t) - F_0(t)}{F_1(t)} = 1 - \frac{F_0(t)}{F_1(t)}$$

Where $F_1(t)$ is the observed cumulative incidence function and $F_0(t)$ the counterfactual cumulative incidence function under the no infection path. To get an estimate of the variance of this estimator we can apply the delta method. We take the ln because the delta method assumes a

normal distribution.

$$\ln\left(1-\frac{F_0(t)}{F_1(t)}\right) \Rightarrow \ln\left(1-\frac{p}{q}\right)$$

We obtain the variance as:

$$\text{var}(g(\hat{p},\hat{q})) = \left(\frac{\partial g}{\partial(p,q)}\right) \text{var}(\hat{p},\hat{q}) \left(\frac{\partial g}{\partial(p,q)}\right)^T$$

$$\frac{\partial g}{\partial p} = \frac{-1}{q-p}$$

$$\frac{\partial g}{\partial q} = \frac{p}{q(q-p)}$$

In g(p,q), p = β and q = logit(F(t)) where F(t) is the observed cumulative incidence function. We know that:

$$F(t) = 1 - \exp(-H(t))$$

so

$$g(p,q) = ln\left(1 - \frac{\left(1 - e^{-\sum_{s=1}^{t} \text{expit}(pX(s))}\right)}{\text{expit}(q)}\right)$$

$$\frac{\partial g}{\partial p} = \frac{-e^{-\sum_{s=1}^{t} \text{expit}(pX(s))}}{\text{expit}(q) - \left(1 - e^{-\sum_{s=1}^{t} \text{expit}(pX(s))}\right)} \times \sum_{s=1}^{t} \text{expit}(pX(s))(1 - \text{expit}(pX(s)))X(s)$$

$$\frac{\partial g}{\partial q} = \frac{-(1 - \text{expit}(q))\left(1 - e^{-\sum_{s=1}^{t} \text{expit}(pX(s))}\right)}{\text{expit}(q) - \left(1 - e^{-\sum_{s=1}^{t} \text{expit}(pX(s))}\right)}$$

We obtain var$(\hat{p}, \hat{q})$ via a sandwich estimator:

$$\frac{1}{n} E^{-1} \left(\frac{\partial U}{\partial (p, q)} \right) \text{var}(U) E^{-1} \left(\frac{\partial U}{\partial (p, q)} \right)^T$$

where $U = \begin{pmatrix} U_p \\ U_q \end{pmatrix}$.

$$\begin{aligned} U_p &= U_{it\bar{a}_t k} \left(h_{t\bar{a}_t k}; \beta \right) W_{i\bar{a}t} \equiv I(\epsilon_{i\bar{a}}(t-1) \neq k) h_{t\bar{a}_t k}(L_{i0}) X W_{i\bar{a}t} \\ &\quad \times \left[I(\epsilon_{\bar{a}}(t) = k) - \text{expit} \left\{ \beta X \right\} \right] \\ U_q &= I(\epsilon_i(t) = 1) - \text{expit}(q) \end{aligned}$$

Partial derivatives U_p:

$$\begin{aligned} \frac{\partial U_p}{\partial p} &= -I(\epsilon_{i\bar{a}}(t-1) \neq k) h_{t\bar{a}_t k}(L_{i0}) W_{i\bar{a}t} X^2 \text{expit}(\beta X)(1 - \text{expit}(\beta X)) \\ \frac{\partial U_p}{\partial q} &= 0 \end{aligned}$$

Partial derivatives U_q:

$$\begin{aligned} \frac{\partial U_q}{\partial p} &= 0 \\ \frac{\partial U_q}{\partial q} &= -\text{expit}(q)(1 - \text{expit}(q)) \end{aligned}$$

Implementation in standard software

Here, we give a step by step procedure and provide syntax for the implementation of the developed methodology in R. In general the main steps in estimating the attributable mortality are:

1. At each time, *select* those patients whose observed data are *compatible with, or, informative about* the hypothetical infection path $\overline{a}$ at that time.

2. Estimate the counterfactual CIF of death at each time t, based on the selected patients at that time and use inverse probability of treatment weighting (IPTW) to account for the selective nature of those patients.

3. Calculate the PAF as the difference between the observed ICU-mortality ($F_1(t)$) and the ICU-mortality that would have been observed (counterfactual) for the same population if infection were prevented for all ($F_{\overline{0}1}(t)$), divided by the observed ICU-mortality.

Data preparation

All available information is stored into two datasets. The `patient` dataset has one line per patient and contains patient characteristics as age, gender, length of ICU-stay, discharged or not, mortality status, infection related information (ever infected, time of infection, antimicrobial) together with the severity of illness indicators measured upon ICU admission. The `followup` dataset contains for every patient as many lines as the ICU length of stay (LOS) (from ICU-admission until ICU-death or discharge). Additional to the information of the `patient` dataset it contains daily measurements of severity of illness indicators.

The first step in the analysis of the data requires several data preparation steps because a patient 's observed data can be compatible with (and thus carry information about) multiple infection paths. Moreover, in the estimation of the subdistribution hazard, patients experiencing the competing event stay in the risk set until the end of follow up. For patients who get discharged

this requires some data expansion until the end of follow up. Patients who get discharged without infection stay compatible with all hypothetical infection paths (see Figure 2.2 showing the compatible infection paths for a patient who gets discharged without infection 20 days after admission (endpoint is 30 day ICU mortality))

The primary data preparation requires following steps:

1. Define new variables (e.g. number of days since start of the intervention, lagged variables,...)

2. Delete the first two days from the dataset because all patients stayed in the ICU for more then 48h and therefore the analysis starts on day 3.

3. Sort the data on the *outcome* and *infection status*.

This last step divides the patients in the dataset into three blocks. First those patients for whom data expansion is not needed (those who died or stayed longer then the follow up time) followed by those for whom the data needs to be expanded with one infection path (patients who were discharged with infection) and finally those for whom the data needs to be expanded with all compatible infection paths (patients who were discharged without infection).

Calculation of the weights

The next step in the analysis involves fitting several models necessary for the calculation of the (stabilized) weights. Therefore, at every day a patient was not previously infected (`inf_1=0`), we fit a model (`M1`) for the daily probability of acquiring infection (`inf`) conditional on baseline characteristics (`base`) and daily measured severity of illness indicators (`dmsi`). The underlying time evolution (`time`) was modeled using cubic splines.

```
M1= glm(inf ~ time + base + dmsi, family=binomial,
data=followup,subset=inf_1==0)
```

```
P1=predict(M1,type = c("response"))
P1=ifelse(inf_1==1,1,P1)
```

For the calculation of the stable weights we need two additional models. One for the daily probability of acquiring infection conditional on the baseline characteristics (`M2`)

```
M2= glm(inf ~ time + base, family=binomial,data=followup,
subset=inf_1==0)

P2=predict(M2,type = c("response"))
P2=ifelse(inf_1==1,1,P2)
```

and a model (`M3`) for the discharge-probability (`disch`) .

```
M3= glm(disch~ time + inf, family=binomial, data=followup)

P3=predict(M3,type = c("response"))
```

In the calculation of the stabilized weights on day t we need the discharge probability on day $t-1$ and therefore define

```
Pgist=c()
Pgist[1]=0

for (i in 2: dim(followup)[1]){
Pgist[i]=ifelse(time[i]==1,0,P3[i-1])
}
```

In a next step, the following algorithm was used to calculate $W_{i\bar{a}t}$ (`W`) and $h_{t\bar{a}_t k}W_{i\bar{a}t}$ (`Wstab`).

```
W=c()
```

```
Wstab=c()
h=c()

for (k in 1:n){ #n=number of patients
tmp=subset(followup,ID==k)
K1=c()
K2=c()
K3=c()
W.tmp=c()
Wstab.tmp=c()
h.tmp=c()
K1=cumprod((tmp$inf*tmp$P1)+((1-tmp$inf)*
      (1-tmp$P1)))
K2=c(1,cumprod((tmp$inf*tmp$P2)+((1-tmp$inf)*
      (1-tmp$P2))))
K3=c(1,cumprod(1-tmp$Pgist))
W.tmp=1/K2
h.tmp=1/(cumsum(tmp$Pgist*(K3[-length(K3)]/K1[-length(K1)]))+
      (K3[-1]/K1[-1]))
Wstab.tmp=d.tmp*W.tmp
W=c(W,W.tmp)
h=c(h,h.tmp)
Wstab=c(Wstab,Wstab.tmp)

cat("iteration=",k, "\n")
}
```

Data expansion

Patients who are discharged (and for whom data expansion is needed) stay compatible with all infection paths so that no selection is made in this subgroup. For those patients, the denominator probability in the weights is equal to 1 and the calculated (stable) weights are constant (equal to the last calculated weight in the observed dataset). The $h_{t\bar{a}_t k}$ part of the stabilized weights only depends on the time of infection (`dayinf`) and are stored in the vectors H_0 and H.

```
H_0=subset(followup,Sort==9)$d


tmp=c()
H=c()
for (i in 1:28){
tmp[i]=unique(subset(followup,dayinf==i)$ID)[1]
H[i]=subset(followup,ID==tmp[i])$h[i]
}
H=ifelse(is.na(H),0,H)
```

Note that in the analysis we were interested in the effect of an infection acquired within 30 days after ICU admission. The 28 in the for loop comes from 30-2 (because the analysis started on day 3).

Because in the data expansion we need information at the end of follow up we construct a new dataset `last` containing per patient the last observation taken at the end of follow up (`LOS_ICU`).

```
last=subset(followup,time==LOS_ICU)
```

Remember that in the first data preparation step the data were sorted on their *outcome* and *infection status* resulting in three blocks of patients. For the actual data expansion we use a for

loop running over each patient's data. By grouping the patients into blocks the algorithm can be fastened. The total number of patients is defined as `tot`. From 1 to `not` no data expansion is needed because these are the patients who died in the ICU or stayed until the end of follow up. From `not+1` to `well` the patients data is expanded until the end of follow up but with only one possible infection regime (*infected on everyday since discharge*). Finally from `well+1` to `tot` full expansion with all compatible infection regimes is done.

To further fasten the algorithm, all compatible infection regimes for a patient discharged at a given day since admission are constructed separately by

```
for (t in 1:27){
Day=c()
A=c()
first=c()
h=c()
Wstab=c()
l=28-t
for (j in 1:t){
tmp=rep(l+j,j+1)
Day=c(Day,tmp)
tmp1=c(0,rep(1,j))
A=c(A,tmp1)
tmp2=c(0,(l+1):(l+j))
first=c(first,tmp2)
tmp8=c(H_0[j],H[(l+1):(l+j)])
h=c(h,tmp8)
}
infdays=A*((Day-first)+1)
```

```
inf=cbind(Day,A,first,infdays,h)
assign(paste("inf",t, sep=""),inf)
}
```

The actual data expansion is done as follows:

```
expand1=c()
for (i in (not+1):well){
l=28-followup$LOS_ICU[i]
Day=(followup$LOS_ICU[i]+1):28
A=rep(1,l)
first=rep(last$dayvap[i],l)
infdays=A*((Day-first)+1)
tmp=cbind(i,Day,A,first,infdays,0,rep(last$h[i]),
    rep(last$W[i],l),rep(last$Wstab[i],l))
expand1=rbind(expand1,tmp)
}

expand2=c()
for (i in (well+1):tot){
l=28-followup$LOS_ICU[i]
inf=get(paste("inf",l, sep=""))[,-5]
d=get(paste("inf",l, sep=""))[,5]
tmp3=cbind(i,inf,0,h,last$W[i],h*last$W[i])
expand2=rbind(expand2,tmp3)
cat("iteration=",i, "\n")
}

expand1=as.data.frame(expand1)
```

```
expand2=as.data.frame(expand2)


names(expand1)=names(followup)
names(expand1)=names(followup)


final_data=rbind(followup,expand1,expand2)


write.csv(final_data,"final.csv",row.names=FALSE,na=".")
```

An independence GEE approach is now used to fit the final analysis model (pooled logistic regression model with patient specific, time varying weights). Because of the data expansion, the `final_data` dataset is to big to use `geeglm` in R (problems with memory allocation). For that reason we did the final analysis in SAS using `proc genmod`.

```
proc genmod data=final_data desc;
class ID;
model ICUmort = time infdays/ dist=binomial;
repeated subject=ID/type=ind covb;
scwgt Wstab;
run;
```

Estimating the attributable mortality

After the analysis in SAS we return to R for the calculation of the population attributable fraction (`PAF`). The estimated parameters from the subdistribution hazard model are stored in `parms_no`. The matrix `XDAT` contains the variables `time` (for 1 to end of follow up) and `infdays` (which is equal to 0 because we want to estimate the `CIF_no` which is the CIF under no infection)

```
p_no=expit(parms_no%*%XDAT)
```

```
CIF_no=1-exp(-cumsum(p_no))
PAF = 1-CIF_no/CIF_obs
```

where `CIF_no` is the counterfactual CIF under no infection and `CIF_obs` is the observed CIF (which can be parametrically or non-parametrically estimated).

The variance of the PAF is obtained in several steps (see the appendix for the formulas). First, for every type of patient we calculate U_p and it derivatives.

```
Up=matrix(nrow=tot,ncol=length(parms_no))
dUp=matrix(nrow=length(parms_no),ncol=length(parms_no),data=0)

for (i in (1:tot)) {
dat=subset(followup,ID==i)
Xdat=cbind(1,time,infdays)
Up.tmp=Xdat*as.numeric(dat$ICUmort-expit(parms_no %*%
       t(Xdat)))*dat$Wstab
dUp.tmp=t(Xdat*as.numeric(expit(parms_no %*%
        t(Xdat))*(1-expit(parms_no %*% t(Xdat)))*
        dat$Wstab))%*%Xdat
Up[i,]=apply(Up.tmp,2,sum)
dUp=dUp+dUp.tmp
}

for (i in (not+1):well){
l=28-patient$LOS_ICU[i]
Day=(patient$LOS_ICU[i]+1):28
A=rep(1,l)
first=rep(last$dayinf[i],l)
```

```
infdays=A*((Day-first)+1)
Xdat=cbind(1,Day,infdays)
Up.tmp=Xdat*as.numeric(0-expit(parms_no %*%
       t(Xdat)))*rep(last$Wstab[i],l)
dUp.tmp=t(Xdat*as.numeric(expit(parms_no %*%
        t(Xdat))*(1-expit(parms_no %*% t(Xdat)))*
        rep(last$Wstab[i],l)))%*%Xdat
u=apply(Up.tmp,2,sum)
Up[i,]=Up[i,]+u
dUp=dUp+dUp.tmp
}

for (i in (well+1):tot){
l=28-patient$LOS_ICU[i]
inf=get(paste("inf",l, sep=""))[,-5]
h=get(paste("inf",l, sep=""))[,5]
Xdat=cbind(1,inf[,1],inf[,4])
Uq.tmp=Xdat*as.numeric(0-expit(parms_no %*%
       t(Xdat)))*(h*last$W[i])
dUq.tmp1=t(Xdat*as.numeric(expit(parms_no %*%
         t(Xdat))*(1-expit(parms_no %*%
         t(Xdat)))*(h*last$W[i])))%*%Xdat
u=apply(tmp,2,sum)
Up[i,]=Up[i,]+u
dUp=dUp+tmp1
}
```

```
#check, needs to be equal to 0 !
apply(Up,2,mean)
```

In a second step we estimate the observed CIF (same principle as for the counterfactual CIF but based on the observed data and without infection in the analysis model) and calculate U_q and its derivatives.

```
Uq=matrix(nrow=tot,ncol=length(parms_obs),data=0)
dUq=matrix(nrow=length(parms_obs),ncol=length(parms_obs),data=0)

for (i in (1:tot)) {
dat=subset(followup,ID==i)
Xdat1=cbind(1,time)
Uq[i,]=t(Xdat1)%*%t(dat$ICUmort-expit(parms %*% t(Xdat1)))
dUq.tmp=t(Xdat1*as.numeric(expit(parms_obs %*%
        t(Xdat1))*(1-expit(parms_obs %*% t(Xdat1)))))%*%Xdat1
dUq=dUq+dUq.tmp
}

for (i in (not+1):tot){
l=28-patient$LOS_ICU[i]
Day=(patient$LOS_ICU[i]+1):28
Xdat1=cbind(1,Day)
Uq.tmp=t(Xdat1)%*%t(-expit(parms_obs %*% t(Xdat1)))
Uq[i,]=Uq[i,]+t(Uq.tmp)
dUq.tmp=t(Xdat1*as.numeric(expit(parms_obs %*%
        t(Xdat1))*(1-expit(parms_obs %*% t(Xdat1)))))%*%Xdat1
```

```
dUq=dUq+dUq.tmp
}

#check, needs to be equal to 0 !
apply(Uq,2,mean)
```

In the next step we calculate the sandwich estimator of the variance as

```
U=cbind(Up,Uq)
VAR_U=var(U)

D11=dUp
D12=matrix(ncol=length(parms_obs),nrow=5,data=0)
D21=matrix(ncol=length(parms_no),nrow=length(parms_no),data=0)
D22=dUq

X <- rbind(cbind(D11/n,D12/n),cbind(D21/n,D22/n))
COV_sand=(solve(X)%*%VAR_U%*%t(solve(X)))/n
```

We use it in the expression obtained from the delta method (see appendix) to calculate the variance of the PAF

```
Xs=cbind(1,time,infdays)
Xs1=cbind(1,time)

lower=c()
upper=c()

for (k in 1:58){
x=as.numeric(expit(parms_no%*%t(Xs))*
```

```
  (1-expit(parms_no%*%t(Xs))))*Xs
y=apply(x,2,cumsum)[k,]
x1=as.numeric(expit(parms_obs%*%t(Xs1))*
   (1-expit(parms_obs%*%t(Xs1))))*Xs1
y1=apply(x1,2,cumsum)[k,]
dgp=-(exp(-cumsum(expit(parms%*%t(Xs))))[k]/CIFno_parm[k])*y
dgq=(CIFobs1[k]/CIFno_parm[k])*(CIFno_parm[k]/CIFobs1[k]^2)*y1
dG=c(as.vector(dgp),dgq)
V=t(dG)%*%COV_sand%*%t(t(dG))
lower[k]=1-exp(log(CIFno_parm[k]/CIFobs1[k])+1.96*sqrt(diag(V)))
upper[k]=1-exp(log(CIFno_parm[k]/CIFobs1[k])-1.96*sqrt(diag(V)))
}
```

3

Attributable mortality of ventilator associated pneumonia: *A reappraisal using causal analysis*.

This chapter is adapted from *Bekaert et al. (2011)* and is the result from a close collaboration between statisticians and physicians in an attempt to apply our new methodology into the ICU-literature and try to convince ICU-experts that the use of more sophisticated method is necessary to answer the question if patients die *from* or *with* infection. With the help of my promoter and a group of ICU-experts from the University Hospital in Ghent (P. Depuydt, J. Decruyenaere and

my co-promotor D. Benoit) we found a good balance between the methodological and medical relevant messages. The work is published in *The American Journal of Respiratory and Critical Care Medicine*. The list of coauthors additionally includes physicians and researchers from the Outcomerea® group who collected the data and helped with the interpretation of the results. (J-F. Timsit, A. Vsin, M. Garrouste-Orgeas, C. Clec'h, E. Azoulay).

Summary

Measuring the attributable mortality of ventilator-associated pneumonia (VAP) is challenging and prone to different forms of bias. Studies addressing this issue have produced variable and controversial results. We estimate the attributable mortality of VAP in a large multicenter cohort using statistical methods from the field of causal inference. Patients (n=4479) from the longitudinal prospective (1997-2008) French multicenter Outcomerea® database were included if they stayed in the ICU for at least 2 days and received mechanical ventilation (MV) within 48 hrs after ICU-admission. A competing risk survival analysis, treating ICU-discharge as a competing risk for ICU-mortality, was conducted using a marginal structural modeling approach to adjust for time-varying confounding by disease severity. 685 (15.3%) patients acquired at least one episode of VAP. We estimated that 4.4% (95% CI 1.6%-7.0%) of the deaths in the ICU on day 30 and 5.9% (95% CI 2.5%-9.1%) on day 60 are attributable to VAP. With an observed ICU-mortality of 23.3% on day 30 and 25.6% on day 60 this corresponds to an ICU-mortality attributable to VAP of about 1% on day 30 and 1.5% on day 60. Our study on the attributable mortality of VAP is the first which simultaneously accounts for the time of acquiring VAP, for informative loss to follow-up after ICU-discharge, and for the existence of complex feedback relations between VAP and the evolution of disease severity. In contrast to the majority of previous reports we detected a relatively limited attributable ICU-mortality of VAP.

3.1 Introduction

Ventilator-associated pneumonia (VAP) is the leading nosocomial infection in mechanically ventilated, critically ill patients, and is commonly considered as a partly preventable disease with a high risk for adverse outcome (Safdar et al., 2005; Bonten et al., 2004). Assessment of the attributable mortality of VAP nevertheless remains challenging (Carlet, 2001; Melsen et al., 2009; Muscedere, 2009). The large variability (Safdar et al., 2005; Muscedere, 2009; Nguile-Makao et al., 2010) in estimated excess risk of death from VAP (between 0 and 70%) has been primarily explained by differences in the patient population under study (case-mix), as well as by the absence of a reference standard diagnosis for VAP, which is usually replaced by a clinical probability coupled with quantitative microbiological data to improve specificity (Timsit et al., 1996b; Koulenti et al., 2009). In our opinion, also the use of various definitions of excess risk contributes to that variation, as well as methodological shortcomings on the level of the data analysis, which may lead to over-or underestimation of the mortality attributable to VAP. As investigators have to rely on observational data alone, careful adjustment for confounding by severity of illness is required to disentangle the complex relationship between VAP and mortality. This is because VAP is essentially a complication of underlying critical illness (Krueger et al., 2002; Rello et al., 1997) and because patients who acquire VAP tend to be more severely ill than patients who do not.

In order to obtain unbiased estimates of the attributable mortality of nosocomial infections, methods should account for obvious findings: (a) patients need to survive long enough in order to acquire infection; (b) patients who acquire infection tend to be more severely ill in the course of their critical illness and not only upon admission; (c) there is a dynamic interplay between VAP and the patients' severity of illness, clinical characteristics and treatment over time; (d) patients who get discharged from the ICU and whose survival time is therefore censored, tend to be in different health conditions as compared to patients staying at the ICU.

This study is the first to estimate the population attributable risk of ICU mortality by VAP in a large size, high quality, multicenter database (Nguile-Makao et al., 2010), whilst overcoming all of the aforementioned obstacles. This will be achieved through the use of statistical techniques from the field of causal inference (Robins et al., 2000; Bekaert et al., 2010b). Similar causal analyses have proved successful in the re-analysis of the *Women's Health Initiative Study* (Hernán et al., 2008; Prentice, 2007) and of epidemiologic studies of AIDS therapies (Hernán et al., 2000) where standard statistical methods contradicted results from the analysis of randomized trials. We adopted the same modeling approach as in these studies, but extended with further refinements which carefully take into account that censoring of the survival time due to discharge from the ICU contains information on the actual survival time (Bekaert et al., 2010b).

3.2 Methods

3.2.1 Study population and data collection

The analysis was based on all records in the longitudinal (1997-2008) French multicenter Outcomerea® database from patients who stayed in the ICU for at least 2 days and received mechanical ventilation (MV) within 48 hrs after ICU admission. In accordance with previous reports (Nguile-Makao et al., 2010), VAP was defined as persistent pulmonary infiltrates on chest radiographs combined with purulent tracheal secretions, and/or body temperature $\geq$ 38.5°C or $\leq$36.5°C, and/or peripheral blood leukocyte count $\geq$10 x 10^9 /L or $\leq$ 4 x 10^9/L; a definitive diagnosis of VAP required microbiological confirmation by quantitative culture from a protected specimen brush ($\geq 10^3$cfu/ml), plugged telescopic catheter specimen ($\geq 10^3$cfu/ml), broncho-alveolar lavage (BAL) fluid specimen ($\geq 10^4$cfu/ml), or endotracheal aspirate ($\geq 10^5$cfu/ml). The effect of the first acquired microbiological proven VAP was modeled.

Data were collected as described previously (Nguile-Makao et al., 2010). In short, the par-

ticipating ICUs provided a random sample of at least 50 ICU-admissions of >24h per year; per case, data entered in the case report form included admission characteristics as well as events and scores following ICU-admission that were recorded on a daily basis. Measured characteristics upon admission consisted of demographic data, admission diagnosis and admission category, chronic illness and comorbidity (using the Knaus definition and including the McCabe score), clinical findings and laboratory investigations. Upon admission and subsequently on a daily basis, the following scores were calculated: Simplified Acute Physiology Score (SAPS) II (Legall et al., 1993), Sequential Organ Failure Assessment (SOFA) (Vincent et al., 1996) and Logistic Organ Dysfunction (LOD) score (Legall et al., 1993). In addition data on daily interventions and treatments were collected: antibiotic treatment, enteral feeding, corticosteroids >0.5mg/kg, invasive or non-invasive mechanical ventilation, vasopressor use, hemodialysis, placement and presence of invasive devices (arterial catheter, central venous catheter, Swan-Ganz catheter and Foley catheter), tracheotomy and do-not-resuscitate (DNR) orders. Throughout this chapter we refer to the above described variables as severity of illness indicators. Antimicrobial treatment was considered immediately appropriate if at the day of microbiological sampling, the patient received at least one antibiotic to which the recovered pathogen(s) was susceptible in vitro.

3.2.2 Statistical analysis

A key challenge for statistical analysis is that VAP and non-VAP patients are inherently different in severity-of-illness (prior to the acquisition of VAP) so that mortality differences between these groups cannot be fully attributed to VAP. When differences in disease severity between these groups are entirely explainable in terms of patient characteristics measured upon ICU admission, then adjustment is possible by including these as covariates in the analysis. This is no longer sufficient when evolution in disease severity contributes as well to the difference between these patient populations. In that case standard regression adjustment for the evolution in severity of illness eliminates the effects of early VAP that are mediated through severity of illness, and in addition induces a so-called collider-stratification bias (Robins et al., 2000; Hernán

et al., 2000; Bekaert et al., 2010a). In view of this, we opted for a marginal structural modeling approach (Robins et al., 2000; Bekaert et al., 2010b), which enables to assess what the ICU-mortality would have been, if all patients would have remained VAP-free or , alternatively, would have acquired VAP on a specific day. Because information on this is lacking for patients who acquired VAP, the observed data for VAP-free patients will be reweighted to predict what the ICU-mortality status would have been for VAP patients, had VAP been prevented for them. For instance, let us suppose that a VAP-free patient, based on his evolution in disease severity, has a 1/3 chance of not acquiring VAP. Then for every such patient, one expects to find 2 patients in the population who experienced a similar evolution in severity-of-illness, but who did acquire VAP. To estimate what the ICU-mortality would have been, if all patients would have remained VAP-free, the data for that VAP-free patient will therefore be counted 3 times, one time to represent himself, and 2 additional times to represent those like patients who did acquire infection. By repeating this for all VAP-free patients, one can thus reconstruct how the data on ICU-mortality status would have looked like, had VAP been prevented for all. The intuitive principle underlying this estimation procedure is illustrated by numerical example in (Robins et al., 2000). The details and validity of the approach in this context are reported in (Bekaert et al., 2010b); we here outline the main steps.

In a first step, we fitted a logistic regression model for the daily probability of acquiring VAP. Because mechanical ventilation could be stopped or restarted on each day, this model includes the patient's time-dependent ventilation status. The obtained probabilities then served to generate a daily patient-specific weighing factor, which was defined as the reciprocal of the probability of this patient having his observed VAP-status and previous history by that day (Bekaert et al., 2010b). Because severity of illness indicators measured on a given day may have been influenced by infection acquired on that day, the model for VAP included only lagged values from the day before. For the SOFA score and antibiotic use we adjusted for lagged values three days (72hrs) before to recognize that the SOFA score and the antibiotic use within 48hrs before the

onset of VAP are possible surrogate markers for an infection which was incubating, and which may therefore be affected by VAP on the given day. All analyses were additionally adjusted for the ICU-center and the admission year.

In the second stage of the analysis, we accounted for informative censoring of the survival time through a competing risk analysis (Prentice et al., 1978; Satagopan et al., 2004) where we considered discharge from the ICU as a competing risk for ICU-mortality (Resche-Rigon et al., 2006b). Our primary focus was on 60-day ICU-mortality because many VAP deaths occur after the first 30 days of ICU stay, and because the analysis of more distal endpoints, such as overall hospital mortality, requires a more stringent confounding adjustment. Using the proposed weights we evaluated hazard regression models, involving the so-called subdistribution hazard, which measures at each time the instantaneous risk of ICU-death at that time amongst patients who did not die within the ICU before that time (Bekaert et al., 2010b). From these models, we inferred the impact of acquiring VAP in the ICU on the Cumulative Incidence Function (CIF) of ICU-death, which is the probability of dying within the ICU before a given time, as a function of time. This was then used to estimate the attributable mortality of VAP as the population attributable fraction (Resche-Rigon et al., 2006b) of ICU-mortality related to infection. On each day, this was calculated as the difference between the observed ICU-mortality and the ICU-mortality that would have been observed for the same population if VAP were prevented for all, divided by the observed ICU-mortality. It can be interpreted as the percentage of the observed ICU-deaths that could be avoided by preventing VAP, or, as the percentage of the observed ICU-deaths who died because of VAP (Tian and Pearl, 2000).

All models were built using a stepwise model selection approach at the 5% significance level using SAS 9.2 (SAS Institute Inc., Cary, NC, USA) and R (R foundation for Statistical Computing, Vienna, Austria).

3.3 Results

3.3.1 Descriptives

4479 patients from the Outcomerea database fulfilled the inclusion criteria, 685 (15.3%) of whom developed pneumonia within 30 days after ICU admission. Forty-one patients (5.9%) were diagnosed with pneumonia more than >48h after discontinuation of mechanical ventilation; while strictly these patients could not be classified as VAP, though still representing ICU-acquired pneumonia, we choose to retain the term VAP throughout the manuscript instead of 'ICU-acquired pneumonia' as it is the more widely used term. Of the 685 patients who acquired VAP, 405 (59.1%) immediately received appropriate antimicrobial treatment. The most prevalent micro-organisms in the 685 episodes of VAP are detailed in table 3.1. From the 685 VAP episodes, 868 micro-organisms were isolated. The main micro-organisms were *P. aeruginosa* (n=227,26.2%), *S. aureus* (n=133,15.3%) and *E. coli* (n=66, 7.6%). At the onset of VAP, the median SOFA score was 6 (25th-75th percentile; 4-8). The median duration between ICU-admission and the onset of VAP was 8 days (25th-75th percentile; 5-12).

The crude 60-day-ICU-mortality risk of the VAP patients was 33.0% (226/685) compared to 24.3% (921/3794) for those who remained free of VAP. The crude 60-day-ICU-mortality in patients with appropriate versus inappropriate antimicrobial therapy was 31.9% (129/405) and 34.6% (97/280), respectively. Median length of stay in the ICU was 7 days (25th-75th percentile; 4-13) and 22 days (25th-75th percentile; 13-38) in patients without and with VAP, respectively. Patient characteristics upon admission are detailed in table 3.2.

3.3.2 Risk of acquiring VAP

The change in hazard of acquiring VAP with time was modeled in a flexible way allowing for differences in evolution depending on the patient's gender (p=0.01), underlying severity of illness (as measured by the SOFA score upon ICU-admission) (p = 0.02), and on antibiotic treatment

	Patients with VAP: (n=685)	Patients without VAP: (n=3794)
Male gender, n (%)	493 (72.0)	2371 (62.5)
Age, mean (SD)	63.2 (15.5)	62.6 (16.8)
ICU-length of stay, median (Q1 , Q3)	22 (14 , 38)	7 (4 , 13)
Ventilation days, median (Q1 , Q3)	19 (11 , 33)	5 (3 , 10)
SAPS II, mean (SD)	49.9 (16.2)	48.4 (18.3)
Admission category		
Medicine, n (%)	476 (69.5)	2317 (62.5)
Emergency surgery, n (%)	114 (16.6)	811 (21.4)
Scheduled surgery, n (%)	93 (13.6)	657 (17.3)
Main symptoms at ICU-admission		
Shock, n (%)	185 (27.0)	1001 (26.4)
Coma, n (%)	143 (20.9)	884 (23.3)
Acute respiratory failure, n (%)	213 (31.1)	846 (22.3)
Other chronic illnesses		
Hepatic,n (%)	53 (7.7)	253 (6.7)
Cardiovascular, n (%)	96 (14.0)	518 (13.7)
Pulmonary, n (%)	139 (20.3)	617 (16.3)
Renal, n (%)	25 (3.6)	147 (3.9)
Immunosuppression, n (%)	83 (12.1)	444 (11.7)
Crude Mortality rates		
30-day ICU-mortality, n (%)	165 (24.1)	876 (23.1)
60-day ICU-mortality, n (%)	226 (33.0)	921 (24.3)
global ICU-mortality, n (%)	237 (34.6)	937 (24.7)

Table 3.1: Characteristics and crude mortality rates for patients with and without VAP.

Micro-organism	n(%)
Gram positive	244 (28.1)
S. pneumonia	45 (5.2)
S. aureus	
Methicillin-susceptible	84 (9.7)
Methicillin-resistant	49 (5.6)
Coagulase-negative staphylococci	32 (3.7)
Enterococci	7 (0.8)
Streptococcus, other	27 (3.1)
Gram-negative	554 (63.8)
H. influenzae	63 (7.3)
Enterobacteriaceae	
E. coli	66 (7.6)
Klebsiella sp.	38 (4.4)
Enterobacter sp.	37 (4.3)
C. freundii	16 (1.8)
S. marcescens	21 (2.4)
P. mirabilis	14 (1.6)
M. morganii	13 (1.5)
Non-fermenting pathogens	
P. aeruginosa	
Wild type	155 (17.9)
Resistance mechanism	72 (8.3)
Acinetobacter sp.	24 (8.2)
S. maltophilia	35 (4.0)
Other	70 (8.1)

Table 3.2: Overview of the main micro-organisms causing VAP. In total 868 micro-organisms were isolated from the 685 VAP episodes.

upon admission ($p < 0.001$). Mechanical ventilation ($p < 0.001$) and enteral feeding the day before the possible occurrence of VAP were associated with an increased risk of acquiring VAP. The magnitude of the effect of enteral feeding increased with higher SOFA scores 72hrs before the possible occurrence of VAP ($p = 0.02$). The impact of SOFA on the risk of acquiring VAP followed opposite directions depending on the value at admission ($p=0.03$), with an increasing risk for patients with an initial value below 10, and a decreasing risk for patients with a higher initial SOFA. A DNR-order ($p=0.02$) and antibiotic treatment on admission and 72hrs before the occurrence of VAP ($p <0.001$) were associated with a lower risk of VAP. Finally, patients with a urinary tract ($p = 0.04$) and catheter related infection ($p <0.001$) 24hrs before the possible occurrence of VAP were more vulnerable to acquire VAP unless the urinary tract infection was acquired more than 6 days ($p=0.02$) before the occurrence of VAP. Note that these results should not be causally interpreted, as they are merely used to create a statistical population in which VAP on each day is independent of measured daily indicators of disease severity, so as to enable an unbiased assessment of the attributable ICU-mortality of VAP.

3.3.3 Attributable ICU-mortality due to VAP

From the marginal structural modeling analysis we found that the (subdistribution) hazard of ICU-death increased with 2.3% per additional day since the onset of VAP (HR 1.023 with 95%CI 1.011 to 1.034; $p< 0.001$). Corresponding increases in hazard ratios of 2.0% (HR 1.02 with 95%CI 1.007 to 1.034; $p=0.003$) and 2.7% (HR 1.027 with 95%CI 1.017 to 1.047; $p=0.001$) were observed for appropriately versus inappropriately treated VAP, respectively. The effect of VAP on ICU-mortality further depended on the patients' severity of illness upon ICU-admission ($p = 0.01$). Table 3.3 gives an overview of the hazard ratios of ICU-death per additional day since infection. The effect of VAP was the largest for patients with intermediate (28-40) SAPSII scores and attenuated in patients with high (>65) or low (<20) SAPSII scores.

Figure 3.1 displays the effect of VAP on ICU-mortality over time. On day 30 and day 60, the population attributable fraction of ICU mortality due to VAP Figure 3.2 equals 4.4% (95%

SAPS II on admission	Hazard ratio of ICU-death per additional day since infection (95%CI)	P-value
15 (5%)	1.023 (0.980 to 1.068)	0.31
20 (10%)	1.030 (0.997 to 1.063)	0.07
28 (25%)	1.037 (1.018 to 1.056)	<.001
40 (50%)	1.038 (1.025 to 1.052)	<.001
53 (75%)	1.027 (1.013 to 1.041)	<.001
65 (90%)	1.00 (0.989 to 1.022)	0.49
73 (95%)	0.990 (0.960 to 1.010)	0.28
Overall	**1.023 (1.011 to 1.034)**	**<.001**

Table 3.3: Hazard ratios of ICU-death per additional day since infection calculated for patients with different SAPSII scores on admission (different percentiles).

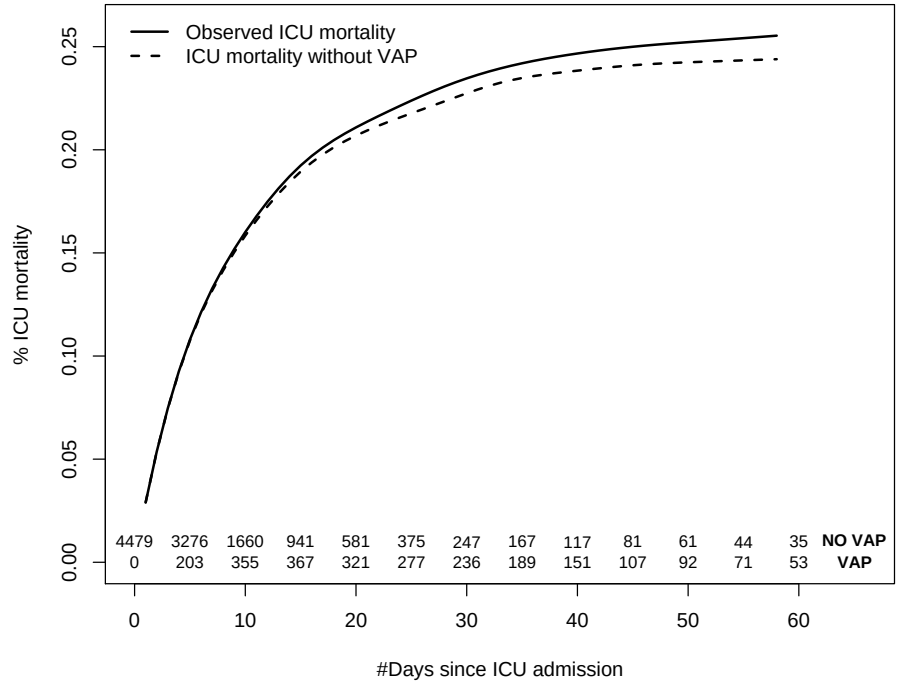

Figure 3.1: The observed cumulative ICU-mortality together with the ICU-mortality as it would have been observed for the same population if VAP were prevented for all.

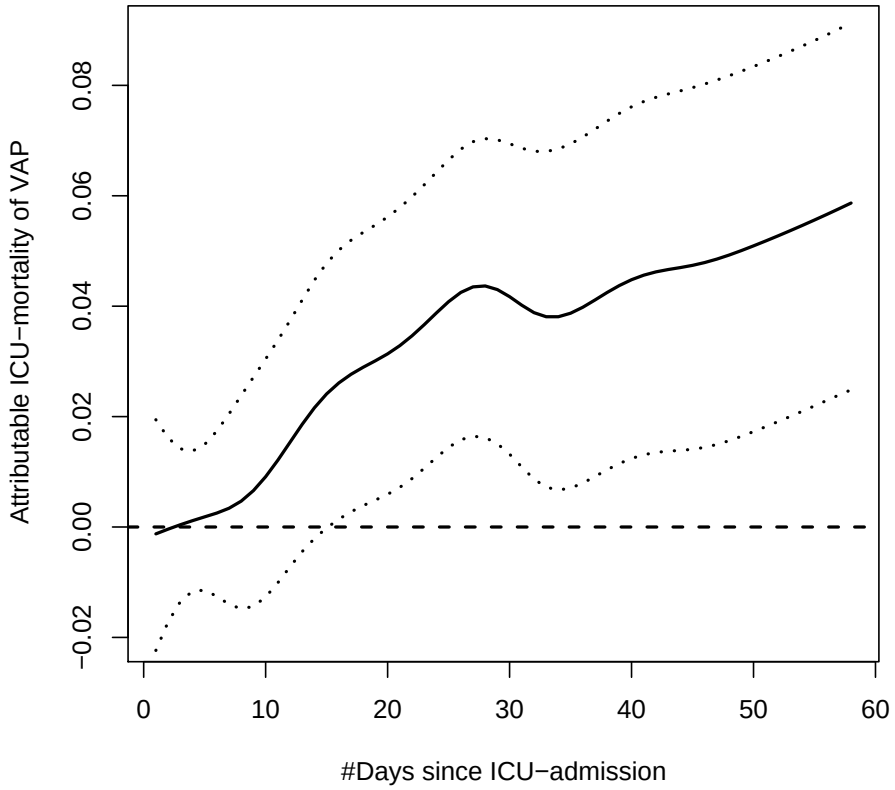

Figure 3.2: The attributable ICU-mortality of VAP as a function of time, defined as the population attributable fraction (PAF). The solid line represents the percentage of ICU mortality that could be attributable to VAP, or, the percentage of the observed ICU-deaths that could be avoided by preventing VAP. The dashed line is the corresponding 95% confidence interval.

CI 1.6%-7.0%) and 5.9% (95% CI 2.5%-9.1%), respectively. For instance, on day 30 this means that if one were able to prevent VAP for all patients in the ICU, 4.4% (95% CI 1.6%-7.0%) of the observed ICU-deaths within the first 30 days following ICU admission could be avoided.

3.4 Discussion

We estimated that 4.4% (95% CI 1.6%-7.0%) of the deaths in the ICU on day 30 and 5.9% (95% CI 2.5%-9.1%) on day 60, could be attributed to VAP, after careful adjustment for time-varying disease severity through the use of a marginal structural modeling analysis, and for ICU-discharge as a competing risk for mortality through the use of competing risk analysis. With an observed ICU-mortality of 23.3% on day 30 and 25.6% on day 60 this corresponds to an attributable ICU-mortality of about 1% on day 30 and 1.5% on day 60. As in previous studies on VAP and hospital acquired bloodstream infections (Nguile-Makao et al., 2010; Buenocavanillas et al., 1994; Kim et al., 2005), our results also indicate that patients who are more critically ill upon admission (higher range of SAPSII) do not experience a major attributable (or additional) effect of VAP compared to those with an intermediate SAPSII score (28-40). The likely explanation for this phenomenon is that the risk of death in patients with established severe organ failure is less modifiable, neither by subsequent treatment nor by intercurrent complications. On the other hand, the lower attributable mortality of VAP in patients in the lower range of SAPSII could be biologically explained by better preservation of host defense mechanisms in these patients (Hotchkiss and Opal, 2010), and by a larger window of opportunity to alter natural history of the infection by appropriate treatment.

The problem of attributable mortality of VAP has already generated a considerable literature. However, only a minority of the studies published thus far performed multivariate analysis to control the association between VAP and mortality for confounding by severity of illness (Melsen et al., 2009; Muscedere, 2009). Furthermore, although a number of authors (Nguile-

Makao et al., 2010; Wolkewitz et al., 2008, 2009; Schumacher et al., 2007) have previously made progress in terms of estimating the attributable effect of VAP by carefully acknowledging the time at which infection is acquired in their analysis, to the best of our knowledge, all previous reports ignored confounding by the evolution of disease severity over time. The challenge is how to incorporate the existence of complex feedback relations between VAP, underlying critical illness and evolving organ failure over time (Bekaert et al., 2010b,a). By using standard statistical regression methods to control for severity of illness indicators, one introduces a so-called collider-stratification bias (Robins et al., 2000; Greenland, 2003). Even when controlling only for severity of illness before infection, such bias results because the severity of illness on a given day is influenced by (the absence of) infection on the previous days. Therefore, whereas the wide range of estimated attributable mortality may reflect true differences resulting from case mix and appropriateness of treatment, in our opinion it additionally reflects incorrect and incomplete adjustment for the condition (and hence the inherent risk of death) of the patient at the very time of developing VAP. The continuous monitoring of vital and biochemical parameters of patients during their stay at the ICU offers the potential to correct for longitudinal information on the patient's health condition evolving over the duration of their critical illness. To incorporate this huge set of data in a meaningful way, a close collaboration between clinicians with a good bedside experience in the treatment of these patients and statisticians who are able to develop and/or use more advanced complex statistical techniques which appropriately account for the aforementioned biases is mandatory. Our study is the first to address all of them via the use of techniques from the field of causal inference.

In contrast to the majority of previous reports (see Melsen et al. (2009) for a systematic review), we obtained a relatively mild attributable ICU-mortality of VAP. A comparison is nevertheless difficult to make because previous studies often focus on alternative risk measures such as relative risks, odds ratios or hazard ratios. Our focus on the population attributable fraction has the advantage that it translates into daily estimates of attributable mortality with a clear medical

interpretation because they express the percentage of observed ICU-deaths by that day whose death was a consequence of VAP. Similarly, they reflect the percentage of ICU-deaths by that day which could be avoided if VAP could be completely prevented.

Since in the popular press, nosocomial infection, including VAP, is increasingly depicted as a preventable disease leading to avoidable death, thus having legal implications, an accurate measure of the attributable mortality of VAP has importance beyond its academic interest. However, it would be erroneous to use our result as a justification to neglect measures to prevent VAP or to minimize the importance of VAP diagnosis and treatment. As in other studies, the attributable mortality of VAP does not reflect the natural history of the disease itself, but the effect of disease modified by therapy, including advanced organ support in addition to antimicrobial therapy. Since modern intensive care can support or partially replace vital functions for prolonged periods of time and for high levels of organ dysfunction, the relationship between VAP and overall mortality is mitigated or obscured. This is congruent to daily clinical practice, in which a direct relationship between occurrence of VAP and death is rarely observed as long as decisions to forego life-sustaining therapy are not taken. Furthermore, similarly to all previous studies, the impact of VAP is measured relative to the absence of VAP, but not to the absence of nosocomial infection altogether, such as urinary tract or catheter-related infections. Finally, our study did not address attributable morbidity of VAP, such as expressed by ICU length-of-stay.

Although our study took into account as much information as possible and used advanced statistical methodology to correct for various biases, several limitations must be acknowledged. First, like all analyses of observational studies, the validity of our analysis relies on the assumption that all relevant confounders have been taken into account. Without confounding adjustment, one can only learn about the association between VAP and mortality. This association is of little account when no causal interpretation can be given, because the relevant scientific question is a causal one: *do patients die from (i.e., due to) or with infection?* Second, our analysis did not

adjust for measurement error in the patient's VAP status. Measurement error of the timing of VAP onset may arise by the process of disease incubation being gradual. This could affect our analysis, which is explicit that cause must precede effect. In view of this, we used lagged values measured from the day before to predict VAP on each day (and of 3 days before in the case of the SOFA score and antibiotic use). It is unclear whether this adjustment sufficiently corrected for an incubation effect, although a sensitivity analysis (not shown) revealed little impact of the chosen lag time. Measurement error may also arise when the VAP status is less recorded for patients with a DNR code, e.g. by refraining from microbiological sampling or chest X-ray. Third, unlike others, we do not provide subgroup analyses. This is because restriction of the analysis to subgroups that are defined on the basis of time-dependent health characteristics may introduce a collider-stratification bias (Bekaert et al., 2010a; Greenland, 2003). For instance, the analysis of subgroups defined by the type of pathogen is challenging because the type of pathogen can be related with time dependent health characteristics. Indeed, it is well know that Pseudomonas aeruginosa often persists in cultures of tracheotomized patients. As such, this pathogen may in a certain number of patients be a surrogate marker of tracheotomy, which by itself is a surrogate marker of severe critical illness polyneuropathy. Further methodological development is needed to enable the assessment of effect modification by time-dependent measures of disease severity. Finally, one should also note that, while all patients involved in this analysis were ventilated for more than 48hrs, 5.9% of the 685 patients who acquired a VAP were already extubated $>$48hrs. This small percentage of patients may not be technically classified as VAP patients. A re-analysis in which these patients are reclassified as non-VAP patients gave similar results: that 3.7% (95% CI 0.5% - 6.8%) of the deaths in the ICU on day 30 and 5.2% (95% CI 1.6%-8.6%) on day 60 are attributable to VAP.

In conclusion, this study on the attributable mortality of VAP is the first which appropriately accounts for the timing of acquiring VAP, for the existence of complex feedback relations between VAP and disease severity, and for informative loss to follow-up after ICU-discharge. In contrast

to the majority of previous reports we detected a relatively limited attributable ICU-mortality.

4

On model selection and model misspecification in causal inference

This chapter is adapted from *Vansteelandt et al. (2010a)* and published in *Statistical methods for medical research* with Gerda Claeskens as a coauthor. My contribution was mainly to the model selection part and less to the part on model misspecification (sections 4.3.2 and 4.3.3).

Summary

Standard variable-selection procedures, primarily developed for the construction of outcome

prediction models, are routinely applied when assessing exposure effects in observational studies. We argue that this tradition is sub-optimal and prone to yield bias in exposure effect estimates as well as their corresponding uncertainty estimates. We weigh the pros and cons of confounder-selection procedures and propose a procedure directly targeting the quality of the exposure effect estimator. We further demonstrate that certain strategies for inferring causal effects have the desirable features (a) of producing (approximately) valid confidence intervals, even when the confounder-selection process is ignored, and (b) of being robust against certain forms of misspecification of the association of confounders with *both* exposure and outcome.

4.1 Introduction

The primary goal of most observational studies is to assess cause-effect relationships. Model-selection procedures - in particular variable-selection procedures - are routinely employed in this process, but rarely with regard to the ultimate focus on causal effects (Hand and Vinciotti, 2003; Claeskens and Hjort, 2003). In addition, the reliance on model-selection procedures is commonly ignored when causal inferences are ultimately drawn. We will reconsider principles of model-selection when the focus is on the estimation of causal effects. We give a brief outline below.

Decisions to exclude/include covariates in a regression model are commonly based on the strength of evidence for their (residual) association with the outcome. When the (causal) effect of a given exposure on the outcome is targeted, then this routine strategy is not ideal and may result in a potentially substantial bias in the exposure effect estimate. The decision to include covariates in a regression model must ideally be based on the strength of evidence for these covariates confounding the association between exposure and outcome. Since by definition, confounders are simultaneously associated with exposure and outcome, procedures that ignore the covariate-exposure association can be sub-optimal, especially for covariates that have strong associations

with the exposure (Crainiceanu et al., 2008; Rubin, 1997). Causal inference procedures that naturally evaluate the strength of covariate-exposure associations (e.g. propensity score adjusted estimators (Rosenbaum and Rubin, 1983)) may thus behave differently than standard (outcome-regression based) procedures, especially when combined with model-selection strategies.

In Section 4.2.1, we argue that the set of potential confounders amongst all measured covariates is often high-dimensional in practice and that there is some tension between the desire to acknowledge all of them through regularization methods, such as ridge regression, and the desire to reduce the covariate space through confounder-selection procedures. We discuss limitations of the most commonly adopted confounder-selection procedures in Section 4.2.3 and argue in Section 4.2.4 that ideally such procedures should directly target the quality of the exposure effect estimator. One proposal is worked out in detail for logistic regression models and applied in Section 4.2.5 to the analysis of an observational study for the effect of right-heart catherization on 180-day mortality in critically ill patients. A limitation to the use of confounder-selection strategies is that they have a tendency to produce under-covering confidence intervals by not acknowledging model uncertainty. In Section 4.2.6 we focus on causal inference procedures that return consistent causal effect estimators when a model for the exposure distribution, given confounders, is correctly specified. We demonstrate that, surprisingly, these procedures remain (approximately) confidence valid in the presence of exposure model selection. For this and other reasons mentioned in the article, they thus succeed better than standard estimation procedures at quantifying the total degree of uncertainty.

In the remainder of this chapter, we focus on the broader problem of model building as opposed to variable-selection. We discuss principles of causal model building in Section 4.3.1 and examine the consequences of model misspecification in Section 4.3.2. In particular, we study misspecification bias affecting so-called doubly robust (Robins and Rotnitzky, 2001) estimation procedures which promise consistent estimation of causal effects when at least one of two

(possibly overlapping) nuisance working models is correctly specified. This leads to estimation procedures that perform well under more global forms of working model misspecification, which are seen to substantially outperform more standard procedures in simulation studies reported in Section 4.3.3.

4.2 Confounder-selection

4.2.1 Confounder-selection versus regularization

Throughout - unless otherwise specified - we assume that a possibly high dimensional collection of covariates is available, which includes all confounders for the effect of exposure A on outcome Y, and thus contains at least one subset of covariates that are sufficient to control for confounding (Greenland et al., 1999a). Determining such subset is impossible in the absence of background knowledge on the causal data-generating mechanism (Robins, 2001). This is largely because adjustment for covariates that are affected by the exposure or the outcome can actually increase bias (Rosenbaum, 1984; Pearl, 2009a; Schisterman et al., 2009), which makes purely associational approaches to confounder selection fallible (Hernan et al., 2002). Causal diagrams (Greenland et al., 1999a; Pearl, 2009a; Robins, 2001) are very helpful to communicate and visualize the data-generating mechanism and, subsequently, to identify covariate sets that are sufficient for confounding control (Greenland et al., 1999a) (see **chapter 1** for an introduction on causal DAG's).

It is presumably true that in most realistic applications, all covariates in a sufficient covariate set will have some association with both the outcome and the exposure (Greenland, 2008). From that perspective, with concern for bias, it seems beneficial to adjust for all available covariates in the set (D'Agostino, 1998; Greenland, 2007, 2008). This has the further advantage that, by acknowledging the uncertainty regarding all covariate effects, it returns a more honest reflection of the overall uncertainty regarding the exposure effect estimator. However, it has the

disadvantage that it may induce a bias and inefficiency as a result of overfitting in the outcome regression model. To guard against this, one could use regularization methods such as ridge regression (see Greenland (2008) and Budtz-Jorgensen et al. (2007) for convincing examples). Alternatively, because propensity-score adjusted estimators can cope better with some overfitting in the propensity score (Robins et al., 1992; Vansteelandt et al., 2010b), one could consider propensity-score adjustment based on a fitted propensity score model which includes all available covariates (D'Agostino, 1998).

The folklore that conditioning on measured covariates reduces bias, must however be taken with caution. This is not only true because of the increased concerns of model misspecification, of possible shrinkage bias and of missing or mismeasured covariate data as more covariates are considered. More fundamentally, evidence is accruing that even adjustment for antecedents of the exposure may induce or aggravate selection bias. This may happen when, as in the causal diagram of Figure 4.1, non-causal relationships are observed between the confounders L and both exposure A and outcome Y. In that case, the adjustment for L induces a so-called M-bias (Greenland et al., 1999a; Greenland, 2003; Pearl, 2009b; Sjolander, 2009) by connecting exposure A and outcome Y along the path $A \leftarrow U1 \rightarrow L \leftarrow U2 \rightarrow Y$. When the causal effects of L on exposure and outcome are weak, this bias may in principle exceed the bias of an unadjusted analysis. In particular, when L affects neither exposure, nor outcome, then interestingly the unadjusted analysis, but not the adjusted analysis, would be valid. A further problem occurs when the association between A and Y is confounded through an unmeasured common cause (i.e., $U3$ in Figure 4.1). In that case, the bias of the unadjusted analysis may surprisingly be amplified upon adjusting for L, provided L is strongly correlated with the exposure (Wooldridge, 2009; Pearl, 2010) (see also Section 4.2.1).

In view of the concerns for M-bias and bias amplification, it may be advantageous to adjust for a strictly smaller subset of covariates that are minimally (Greenland et al., 1999b) sufficient to control for confounding (in the sense that, given these covariates, all remaining covariates are

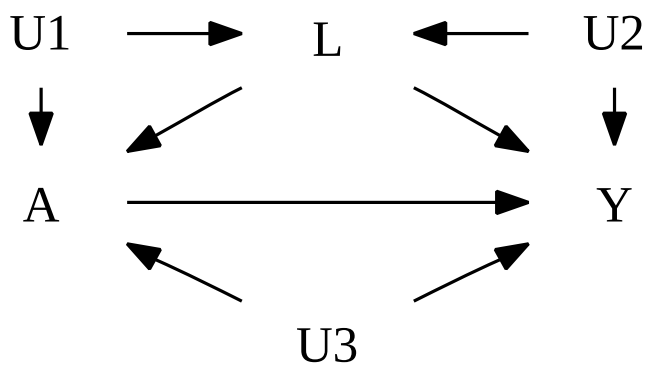

Figure 4.1: Causal diagram with measured variables A, L and Y, and with $U1, U2$ and $U3$ unmeasured variables.

only associated with either the exposure or the outcome, but not both). Adjusting for a subset of available covariates may have the further advantage of yielding more efficient effect estimators. In particular, Hahn (2004) elegantly shows that adjustment for covariates that have no (residual) association with the outcome can reduce the efficiency of nonparametric estimators of the marginal treatment effect, unlike adjustment for covariates that have no (residual) association with the exposure. In view of this and of the aforementioned concerns about bias amplification, it has been suggested that the selection of confounders should be based on their importance with respect to the outcome, rather than the exposure (Pearl, 2010; Brookhart et al., 2006). Whether such recommendation to reduce a sufficient set of confounders is successful, is arguable however. First, the results of Hahn (2004) refer to settings where *a priori* knowledge is available that certain covariates have no residual association with the outcome. In practice, the selection of confounders is virtually always (at least partly) data-driven, but the ensuing uncertainty is most often ignored. Upon acknowledging the additional model uncertainty, one may well find effect estimators obtained after variable-selection being less efficient than those obtained from a full model which includes all available covariates (Budtz-Jorgensen et al., 2007; Claeskens and Hjort, 2008). Second, in the next section we will find that in well designed studies where efforts have been made to collect data on causal risk factors for the exposure that are also associated with the outcome, the concerns for M-bias and bias amplification may be more modest. Third,

even when these concerns are justified, then as a result of multicollinearity, it would still be difficult to measure the importance of a covariate with respect to the outcome whenever that covariate is strongly correlated with the exposure. Standard variable-selection based on hypothesis testing in outcome regression models may therefore lack power to detect even relatively strong confounders.

Extensive simulation studies are needed, complementing the early work of Mickey and Greenland (1989) and Maldonado and Greenland (1993), to be able to gauge the relative importance of the aforementioned pros and cons of confounder-selection versus no selection. Making a choice between these strategies is further complicated by the fact that M-bias and bias amplification occur only in the presence of *unmeasured* common causes of exposure, outcome and confounders, so that one cannot protect against it or know to what extent these biases - which primarily affect strategies that avoid selection - are present. In the following section, which may be skipped by the less interested reader, we therefore attempt to develop insight into the extent to which the concerns for M-bias and bias amplification are justified in practical applications.

4.2.2 M-bias and bias amplification

We compute the magnitude of the biases of the unadjusted and adjusted analysis in the Appendix for multivariate normal variates following the path diagram of Figure 5.1, extending the work of Wooldridge (2009) and Pearl (2010). Let ρ_1 denote the standardized path coefficients (Wright, 1934) between A and $U1$, L and $U1$, L and $U2$, or Y and $U2$ (which we assume to be equal for simplicity), ρ_2 denote the standardized path coefficients between A and $U3$, or Y and $U3$ (which we assume to be equal for simplicity), ρ_{al} denote the correlation between A and L and ρ_{yl} denote the correlation between Y and L in the absence of an exposure effect (or upon setting A to a fixed value, uniformly in the population). Then the bias of the adjusted analysis (either based on standard regression adjustment or based on inverse probability weighting by $1/f(A|L)$; see

Section 4.3.2) is

$$\frac{\rho_2^2 - \rho_1^4}{1 - \rho_{al}^2},$$

where the first term reflects bias due to the unmeasured common cause $U3$ of A and Y, and the second term reflects M-bias; that is, the two terms reflect spurious associations along the paths $A \leftarrow U3 \rightarrow Y$ and $A \leftarrow U1 \rightarrow L \leftarrow U2 \rightarrow Y$, respectively. The denominator suggests that strong correlations between exposure and measured confounders not only have a tendency to amplify bias resulting from unmeasured confounders $U3$, in line with the conclusions of others (Wooldridge, 2009; Pearl, 2010), but also M-bias. The bias of the unadjusted analysis is

$$\rho_{al}\rho_{yl} - \rho_1^4 + \rho_2^2,$$

which does not suffer this amplification. Here, the first two terms encode bias due to not adjusting for the measured confounder L and the last term measures bias due to the unmeasured common cause $U3$ of A and Y; that is, the three terms reflect spurious associations along the paths $A \leftarrow U1 \rightarrow L \rightarrow Y$, $A \leftarrow L \rightarrow Y$, $A \leftarrow L \leftarrow U2 \rightarrow Y$ and $A \leftarrow U3 \rightarrow Y$. We thus find that the adjusted analysis will have larger bias than the unadjusted analysis when the correlation between Y and L (other than through A) is sufficiently weak in the sense that

$$\rho_{yl} < \frac{\rho_{al}}{1 - \rho_{al}^2}(\rho_2^2 - \rho_1^4). \tag{4.1}$$

Even if the presence of an unmeasured common cause $U3$ of A and Y could be ruled out, the existence of unmeasured common causes such as $U1$ and $U2$ would be difficult to exclude in any given application. In particular when L is high-dimensional, it would be difficult to believe that all of its components are only linked to A or Y by means of a causal effect. This suggests that M-bias is likely to arise in practice, although the fourth order terms express that its magnitude is likely going to be small. Similar findings were obtained by (Greenland, 2003) in the all binary case. An exception occurs when the correlation between A and L is strong, for then even a

modest degree of M-bias may in principle be amplified by a potentially important magnitude.

4.2.3 Confounder-selection strategies

Amongst the various confounder-selection strategies that are routinely adopted in practice, backward elimination based on hypothesis tests in outcome regression models is the default strategy. It is not ideal, however, because it is based on accepting the null hypothesis when covariates are non-significantly associated with the outcome (Greenland, 2008) and because it ignores the association between exposure and covariates when deciding whether a given covariate confounds the association between exposure and outcome (Crainiceanu et al., 2008). As such, it has a tendency to under-select important confounders (Greenland and Neutra, 1980) by ignoring covariates that have relatively weak associations with the outcome (conditional on the exposure), but strong associations with the exposure (Crainiceanu et al., 2008; Rubin, 1997). Such covariates are typically dismissed because they induce problems of multicollinearity (arising from correlation between the exposure and covariates), thereby inflating the uncertainty on the estimated treatment effect. This uncertainty is often interpreted as a sign of inefficiency, which is justified in some cases but should more generally be viewed as a reflection of the lack of information about the exposure effect (Tan, 2008). By eliminating these covariates, one thus risks not only to induce a bias in the estimated exposure effect, but also to understate the actual uncertainty. Precisely in settings where there is much separation in the covariate distributions of exposed and unexposed subjects, and therefore much uncertainty about the exposure effect, conventional backward elimination strategies will tend to remove covariates from the outcome regression model and, thereby, yield misleadingly precise exposure effect estimates. Similar concerns apply to penalization methods such as the lasso or elastic nets (Tibshirani, 1996; Zou and Hastie, 2005) and certain confounder-selection methods based on identification results for minimally sufficient sets of confounders (Greenland et al., 1999b; De Luna et al., 2010) because of their tendency to dismiss covariates that are strongly associated with the exposure.

In epidemiology, some of these concerns have contributed to the popularity of change-in-estimate procedures which tend to have better success (Greenland and Neutra, 1980; Mickey and Greenland, 1989; Maldonado and Greenland, 1993) by directly evaluating the impact of confounder-selection on the magnitude of the exposure effect estimate. While these target more directly a reduction of confounding bias, also these approaches are not ideal because they ignore estimation uncertainty and may be inefficient by under-selecting covariates that are only predictive of the response (Greenland, 1986). Furthermore, apart from finite-sample imprecision and model misspecification, inclusion of a covariate in a regression model may induce a change in treatment effect estimate, even when that covariate is not a confounder of the exposure-outcome relation. This may happen as a result of non-collapsibility of association measures in nonlinear models (Mickey and Greenland, 1989; Maldonado and Greenland, 1993; Greenland et al., 1999b), which may change in magnitude upon adjusting for a covariate that is solely associated with the outcome (but independent of the exposure). This may also happen as a result of M-bias or bias amplification in both linear and nonlinear models.

4.2.4 Focused confounder selection

We believe that an 'optimal' confounder-selection strategy should focus on the quality of the exposure effect estimator. We will therefore closely follow the idea of change-in-estimate procedures, but accommodate their limitations, albeit necessarily presupposing that there are no unmeasured confounders (i.e. in particular, that $U3$ and either $U1$ or $U2$ are absent in the causal diagram of Figure 4.1). Specifically, let τ^* denote the target effect parameter and $\hat{\tau}$ an estimator of it. Then we will focus confounder-selection on the precision of the exposure effect estimator, as measured through its mean squared error $E\{(\hat{\tau} - \tau^*)^2\}$. Our choice not to pursue conventional confounder-selection procedures based on the likelihood function (e.g. based on the AIC or BIC), is further guided by the fact that, as shown in Section 4.3.2, standard maximum likelihood inference can be sub-optimal for the estimation of nuisance working models (e.g. for modeling the association of confounders with either the outcome or exposure).

Mean squared error is also the focus of Claeskens and Hjort (2003), whose focused information criterion (FIC) is based on exact or asymptotic calculations in parametric models, and of Brookhart and van der Laan (2006) who use cross-validation instead. Alternatively, one could focus model/confounder selection on the (counterfactual) prediction error, as in Claeskens et al. (2006), who use a prediction-focused information criterion, and Mortimer et al. (2005) and Haight et al. (2010) who use cross-validation instead.

Given our focus on the mean squared error of the exposure effect estimator, an important consideration is whether the estimators $\hat{\tau}_S$ corresponding to different models S are all consistently estimating the same parameter τ^* under correct model specification. This is not usually the case for conditional exposure effects due to noncollapsibility of nonlinear association measures (Greenland et al., 1999b) and the possibility of effect modification. This makes approaches for model-selection focused on the mean squared error not entirely appropriate for estimating the usual conditional exposure effects. This problem can be overcome by targeting confounder-selection at the marginal or population-averaged exposure effect. For instance, let A be a dichotomous exposure (taking values 0 and 1) and consider the parameter β^* indexing $\text{logit}P(Y=1|A,L) = \omega(L;\gamma^*) + \beta^* A$, where is $\omega(L;\gamma)$ a known function, smooth in γ, and γ^* is an unknown finite-dimensional parameter. For instance, $\omega(L_i;\gamma) = \gamma_0 + \gamma_l L_i$ in the case of standard regression adjustment, or $\omega(L_i;\gamma) = \gamma_0 + \gamma_p \pi(L_i;\gamma)$ with $\pi(L_i;\gamma) = P(A_i = 1|L_i;\gamma) = \text{expit}(\gamma_1 + \gamma_l L_i)$ in the case of propensity score adjustment (Rosenbaum and Rubin, 1983). Then, with $Y(a)$ denoting the counterfactual outcome following exposure level a, the marginal causal odds ratio $\tau^* = \text{odds}\{Y(1) = 1\}/\text{odds}\{Y(0) = 1\}$ can, for given estimates $\hat{\gamma}$ of γ^* and $\hat{\beta}$ of β^*, be estimated as

$$\hat{\tau} = \frac{\sum_{i=1}^n \text{expit}\left\{\omega(L_i;\hat{\gamma}) + \hat{\beta}\right\} / \sum_{i=1}^n \text{expit}\left\{-\omega(L_i;\hat{\gamma}) - \hat{\beta}\right\}}{\sum_{i=1}^n \text{expit}\left\{\omega(L_i;\hat{\gamma})\right\} / \sum_{i=1}^n \text{expit}\left\{-\omega(L_i;\hat{\gamma})\right\}}. \tag{4.2}$$

Thus focussing on the marginal treatment effect τ^*, we propose the following focused confounder-selection procedure, which inherits from work by Claeskens et al. (2006) and Crainiceanu et al. (2008). We divide the model space into $M + 1$ orbits, where M is the number of potential covariates (i.e., confounders and/or functions of confounders, such as higher order terms or interactions) and where the jth orbit, $j = 1, \ldots, M+1$ comprises all models with $j-1$ covariates and an intercept. Within each orbit, we select the outcome regression model that minimizes the mean squared error of $\hat{\tau}$. This is done using the following stochastic search method, which is closely linked to that in Crainiceanu et al. (2008). Starting from a model in the $(j-1)$th orbit, we add the covariate that provides the largest reduction in mean squared error. The stochastic search then selects at random one covariate which is in the model and one which is not in the model, and constructs a new model by interchanging both covariates. The new model is accepted when $L_{\text{new}} < L_{\text{old}}$, where L_{old} and L_{new} are the mean squared errors of $\hat{\tau}$ under the old and new model, respectively. When $L_{\text{new}} > L_{\text{old}}$, the new model is accepted with probability $(L_{\text{old}}/L_{\text{new}})^{\alpha}$, where α is a user-selected tuning parameter. Alternatively, a deletion/substitution/addition algorithm (Haight et al., 2010) could be used, which involves exhaustive model search within model subclasses obtained by either deleting, substituting or adding one covariate to those already available in the model. In this process, the mean squared error can be estimated based on a cross-validation procedure where the data are partitioned into a training sample and validation sample V times. That is, the mean squared error of the estimator $\hat{\tau}$ can be approximated with $(1/V)\sum_{v=1}^{V}(\hat{\tau}_v - \hat{\tau}_0)^2$, where $\hat{\tau}_v$ is the estimator of τ^* as obtained under the considered model on the training sample, and where $\hat{\tau}_0$ is an estimator of τ^* as obtained under the full model on the validation sample. Minimization of this estimated loss function is then equivalent to minimization of the mean squared error when the estimator $\hat{\tau}_0$ is unbiased (Brookhart and van der Laan, 2006). Computing time can be drastically reduced through asymptotic approximations of the mean squared error, which can be made under a local misspecification assumption (see Section 4.2.6). A framework for this is developed in Hjort and Claeskens (2003) for parametric models and adapted to our specific setting in the appendix.

4.2.5 Application

We evaluate the proposed confounder-selection procedure in an observational study investigating the effect of right heart catherization (RHC) on 180 day mortality in 5735 critically ill patients (Connors et al., 1996). For every patient, the exposure of interest A was coded 1 if RHC was used within 24 hours of admission and 0 otherwise. In total, 61 covariates (L) on the patients' underlying health condition within 24h of ICU admission (physiological status), on their underlying comorbidity and on demographic information were available for analysis. The original analysis (Connors et al., 1996) used logistic regression to develop an estimated propensity score for each patient, which was then used for matching RHC patients to non-RHC patients. In this Section, we will contrast different confounder-selection methods, including the one proposed in the previous section. R-code for the analyses can be found in the appendix

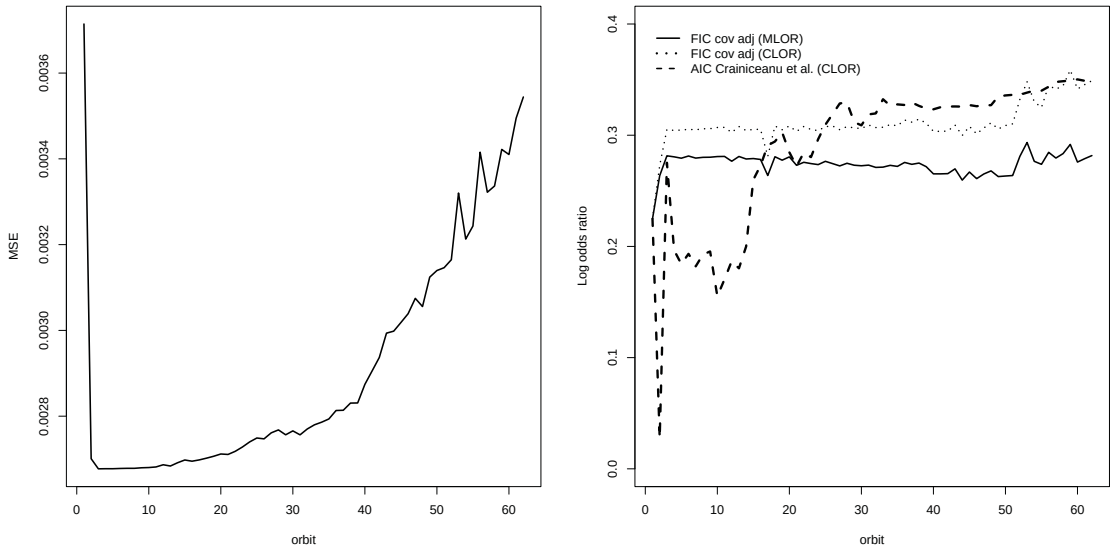

Figure 4.2: Left: Mean squared error (MSE) of the best model within each orbit which is obtained by minimizing the mean squared error of the marginal log odds ratio (MLOR) Right: Estimates of the marginal and conditional log odds ratio as obtained through FIC-based covariate adjustment, and through AIC-based selection as in Crainiceanu et al.

Figure 4.2 (left) shows the mean squared error (MSE) for the best model within each orbit

as obtained by minimizing the mean squared error of the marginal log odds ratio (MLOR) (or equivalently, by minimizing the focused information criterion (FIC) which measures the mean squared error of the MLOR up to an additive constant, see the Appendix) in the case of standard covariate adjustment (i.e., $\omega(L_i;\gamma) = \gamma_0 + \gamma_l L_i$ in Section 4.2.4). The MSE is largest for the narrow (due to large bias) and full model (due to large variance); minimal MSE is attained for simple models involving 2 covariates only. For illustrative purposes, Figure 4.2 (right) compares the thus obtained estimates for the MLOR under standard covariate adjustment (solid black line) with the conditional log odds ratios (CLOR) corresponding to the same models (dotted line). It demonstrates the stability of the estimated MLOR over the different orbits, which is useful information in itself as the observed stability strengthens confidence in the analysis results. It suggests also increasing conditional treatment effect estimates over different orbits, which is due to noncollapsibility of the odds ratio. This underscores the potential limitations of variable-selection procedures that focus on conditional effect measures, which tend to mix confounding with non-collapsibility of association measures.

Figure 4.2 (right) also displays the results obtained upon applying the procedure advocated in Crainiceanu et al. (2008). This procedure involves first selecting covariates on the basis of their association with the exposure as measured in terms of the AIC, and subsequently selecting any remaining covariates on the basis of their residual association with the outcome, again measured in terms of the AIC. The estimates for the CLOR (which is the focus of that procedure) are initially very unstable as a result of selecting covariates that are strongly associated with the exposure, but not with the outcome. Stability in the estimates is attained only for very large orbits which, again, may be partly due to non-collapsibility of the odds ratio. Like standard model selection procedures, it thereby gives a somewhat misleading impression that the association between RHC and mortality is confounded by many of the measured covariates. The proposed procedure improves upon this (a) by focusing the model selection on a parameter which is identically defined over the different orbits, and (b) by selecting covariates on the basis of their

potential to increase the precision of the treatment effect estimate, as well as their ability to reduce confounding in the treatment effect estimate.

Model selection technique	# covariates	COR	MOR	95% CI	MSE	FIC
Unadjusted analysis	0	1.25	1.25	[1.14 to 1.37]	0.0037	5.99
Full model	61	1.42	1.32	[1.18 to 1.49]	0.0035	5.01
BE covariate adjustment	15	1.39	1.31	[1.17 to 1.46]	0.0033	3.87
AIC (Crainiceanu et al.)	47	1.42	1.33	[1.18 to 1.49]	0.0035	4.94
FIC covariate adjustment	2	1.36	1.33	[1.21 to 1.46]	0.0027	0.04

Table 4.1: Estimates of the effect of RHC on mortality, as obtained using different confounder-selection techniques and reported in terms of the conditional odds ratio (COR), the marginal odds ratio (MOR) with 95% confidence interval, MSE (mean squared error) and the FIC (focused information criterion).

Table 4.1 reports estimates of the effect of RHC on mortality as obtained from these different confounder-selection procedures. As shown in Figure 4.2 (left), minimal MSE is attained for simple models involving 2 covariates only (age and a covariate which indicates the presence of a Solid Tumor, Metastatic Disease, Chronic Leukemia/Myeloma, Acute Leukemia or Lymphoma) and results in a MOR of 1.33 (95% CI 1.21 to 1.46). The unadjusted analysis gave a MOR of 1.25 (95% CI 1.14 to 1.38) with an MSE of 0.0037, versus 0.0035 for the full model. In contrast, the 'optimal' model of Crainiceanu et al. (2008) includes 36 predictors of right heart catherization, regardless of their association with the outcome, and 11 additional covariates on the basis of their residual association with the outcome. Also covariate adjustment and propensity score adjustment based on backward elimination (BE) strategies tend to select many more covariates at the expense of accuracy. They do so because the decision to enter covariates into the model is based on either their association with the outcome (as in standard covariate adjustment), or their association with the exposure (as in propensity score adjustment), but not on the basis of a more balanced evaluation in terms of the quality of the treatment effect estimate. Given the large number of patients in this study, many of these associations are strong in terms of the evidence provided by p-values, but not necessarily in terms of their potential to distort the treatment-outcome association by an important magnitude.

4.2.6 Model uncertainty

In small data sets where the variance of the exposure effect estimator is dominant, focused confounder-selection strategies might have a tendency to delete confounders when their adjustment causes a large variance inflation. While this may be beneficial to the overall accuracy of the exposure effect estimator, a concern is that it may come at the expense of confidence validity, considering that confidence intervals capture sampling variability, but not bias. Confidence validity may be further compromised by the fact that uncertainty resulting from the data-driven model building process is commonly ignored. Although the bootstrap or asymptotic approximations (Hjort and Claeskens, 2003) could be used to acknowledge this, these are often not considered in practice. We will now argue that these concerns can be tempered to some extent by the use of propensity-score based estimators.

First, propensity-score based estimators which force important predictors of the exposure into the propensity score (e.g. the procedure advocated by Crainiceanu et al. (2008)) are relatively less susceptible to bias resulting from insufficient confounding adjustment because the set of confounders forms a subset of the exposure predictors; the same is not true for approaches which rely on outcome predictors because the magnitude of the residual association between outcome and predictor, given the exposure, is difficult to assess for predictors that are strongly associated with the exposure. A drawback is that such propensity-score based procedures can be inefficient and more prone to bias amplification when they include predictors that have (almost) no residual association with the outcome (Pearl, 2010; Brookhart et al., 2006).

Second, in the Appendix, we study the asymptotic behaviour of exposure effect estimators which solely rely on correct specification of a propensity score model, as obtained after model-selection. Examples are the G-estimator (Robins et al., 1992) and the inverse probability weighted estimator (Robins et al., 2000) of the average causal effect (see Section 4.3.2). Because the potential for model misspecification cannot be ignored in the presence of model-selection, the

asymptotic behaviour of such estimators is examined within the local misspecification framework of Hjort and Claeskens (2003). More precisely, we assume that the true exposure data-generating mechanism is of the form $f(A|L) = f(A|L; \alpha_1^*, \alpha_2^* + \delta/\sqrt{n})$, with $f(A|L; \alpha_1, \alpha_2)$ a conditional density function of A, given L, which is smooth in α_1 and α_2, where α_1^* and δ are unknown finite-dimensional parameters and where α_2^* is a chosen finite-dimensional parameter (e.g. $\alpha_2^* = 0$). Here, α_1 encodes the unknown part of the parameter vector which is shared between all competing submodels. Each exposure working model, denoted S, thus assumes some of the components α_{-S} of α_2 to be known and equal to the corresponding components of α_2^*, and assumes the remaining components α_S to be unknown. Note that the reason to allow for misspecification of the model parameters within a 1 over root-n distance is because in large samples standard model selection techniques would systematically choose the narrow model (which assumes α_2 equals α_2^*) when smaller misspecifications are considered, and systematically select the wide model (which assumes α_2 is unknown) when larger misspecifications are considered (Hjort and Claeskens, 2003).

In the Appendix, we then show that interestingly a conservative asymptotic variance of the considered exposure effect estimators is obtained when imprecision due to estimation and model-selection on the propensity score is ignored, provided that an efficient estimator is used for the parameters indexing the propensity score model. This result is of importance as it suggests that the model/confounder-selection procedure can be ignored in inference about the exposure effect, provided that the local misspecification assumption holds. It does not immediately follow, however, that confidence intervals which ignore estimation and model uncertainty in the propensity score will attain the nominal coverage probability. This is because, as shown in the Appendix, the distribution of the exposure effect estimator in the presence of model-selection is not centered at zero, but follows a mixture distribution with bias components converging at root-n rate to zero. Preliminary simulation studies (not shown) confirmed that, nonetheless, close to nominal coverage levels are attained even when this is ignored.

4.3 Model building

4.3.1 Principles of causal model building

We will now broaden the focus from confounder-selection to model building. Though historically, the use of parametric models combined with maximum likelihood inference has been dominant (cfr. structural equation models (SEMs)), more recently - stimulated by pioneering work of James Robins - a trend is now seen towards semi-parametric modeling of causal effects. Path diagrams, used by SEM practitioners as convenient representations of a multivariate normal model and as convenient tools for combining path-specific effects into exposure effects of interest, are substituted by 'non-parametric' causal diagrams (Pearl, 1998); these can be combined with semi-parametric models directly parameterizing the exposure effect of interest (Robins, 1997).

The appeal of semi-parametric inference for causal effects surmounts the usual concerns for model misspecification and limited flexibility in parametric inference. Parametric likelihood-based procedures explicitly ignore information on the exposure distribution which has nevertheless demonstrated to be relevant for confounder-selection in Sections 4.2.5 and 4.2.6. For instance, in the absence of an exposure effect, the common strategy of forcing the exposure into the model may lead one to systematically ascribe an effect of extraneous covariates to an exposure effect (Robins and Greenland, 1986). This can be overcome using propensity score methods which force the propensity score into the outcome regression model, irrespective of whether it is significantly associated with the outcome. Robins and Ritov (1997) underscored more formally the importance of using information on the exposure distribution in causal inference by demonstrating that, due to the curse of dimensionality, likelihood-based procedures fail to estimate treatment effects in randomized experiments where randomization is conditional on a high-dimensional covariate; see also (Vansteelandt et al., 2008; Rosenblum and van der Laan, 2009).

In further clarification of the philosophical principles behind semi-parametric modeling of causal effects, suppose that interest lies in the direct effect of A on Y which is not mediated by M in the causal diagram of Figure 4.3.1. SEM procedures would typically dismiss U from the path diagram and thereby arrive at biased causal effect estimates. Alternatively, they would include U, thus requiring models for the conditional densities $f(Y|A, M, U)$, $f(L|U)$ and $f(U)$, and subsequently yield causal effects conditional on U. These are not only difficult to specify, estimate and interpret by the fact that U is unmeasured, but additionally raise questions as to whether the identification of the direct effect under the model comes from structural assumptions (e.g., assumptions about the absence of specific direct effects or common causes) alone, or from parametric assumptions (e.g. regarding the distribution of U) in addition. In the latter case, we say that the considered causal effect is not non-parametrically identified (Scharfstein et al., 1999). This is not ideal as it can make the results heavily sensitive to the chosen (semi-)parametric modeling assumptions (see e.g. Little (1985) and Scharfstein et al. (1999); see Vansteelandt (2009a) for an example illustrating the importance of nonparametric identification in a more general context).

G-computation (Robins, 1986) enables identifying the counterfactual mean $E\{Y(a, m)\}$ corresponding to setting the exposure A at a and the mediator M at m as $\int E(Y|A = a, M = m, L)f(L|A = a)dL$, where the conditional mean $E(Y|A, M, L)$ and density $f(L|A)$ could be substituted with parametric likelihood-based estimators. Also this approach is not ideal as it does not directly parameterize the (controlled) direct effect (Robins and Greenland, 1992) $E\{Y(a, m) - Y(a^*, m)\}$ of interest and henceforth does not enable researchers to express hypotheses of interest (e.g., that a direct effect of A on Y is not modified by M) in a parsimonious way. In addition, it is essentially impossible to postulate nonlinear models for $E(Y|A, M, L)$ and $f(L|A)$, which accommodate a dependence on A (as suggested by Figure 4.3.1), and are such that $\int E(Y|A = a, M = m, L)f(L|A = a)dL$ does not depend on a for all m. This is the

root cause of the so-called null paradox (Robins, 1997) according to which G-computation based tests of the null hypothesis of no direct effect will with certainty be rejected in large samples. These subtleties underscore the importance of parameterizing the exposure effect of interest directly, which may be most naturally approached through the use of semi-parametric inference (Robins, 1999; VanderWeele, 2009; Vansteelandt, 2009b).

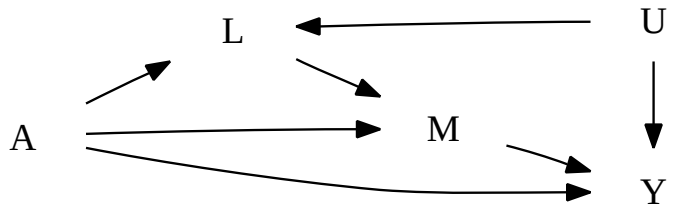

Figure 4.3: Causal diagram with measured variables A, L, M and Y, and with U an unmeasured confounder of the L-Y relationship.

4.3.2 Model misspecification

Many semi-parametric procedures for causal effects separate the modeling of confounders from the modeling of the causal effects of interest. The use of complex confounder models thus need not complicate the interpretation of results; however, their misspecification may induce a bias in the exposure effect estimator. To enrich our understanding, we study the impact of working model misspecification in more detail for so-called G-estimators (Robins et al., 1992) and inverse probability weighted (IPW) estimators (Robins et al., 2000) under the assumption that

$$E(Y|A, L) = \omega_0^*(L) + \tau^* A$$

for some unknown function $\omega_0^*(L)$ of L, where τ^* encodes the exposure effect. For simplicity of exposition, we assume that A is a dichotomous exposure, taking values 0 and 1, and that Y is a continuous outcome. The G-estimator (Robins et al., 1992) is obtained as the solution to an

estimating equation of the form

$$0 = \sum_{i=1}^{n} \{A_i - \hat{\pi}(L_i)\} \{Y_i - \tau A_i - \phi\hat{\omega}_0(L_i)\},$$

where $\hat{\omega}_0(L)$ and $\hat{\pi}(L)$ are estimates of $\omega_0^*(L)$ and the propensity score $\pi^*(L) = P(A = 1|L)$, respectively, based on possibly misspecified models. Further, ϕ is a user-specified constant. If set to 0, it yields the so-called G-estimator which is a consistent and asymptotically normal (CAN) estimator of τ^* if $\hat{\pi}(L)$ is a consistent estimator of $\pi^*(L)$ for all L. In linear models, this estimator is equivalent with the ordinary least squares estimator obtained via regression adjustment for the propensity score (Rosenbaum and Rubin, 1983). If set to 1, it yields the so-called doubly-robust G-estimator which is a CAN estimator of τ^* if for each L, either $\hat{\pi}(L)$ is a consistent estimator of $\pi^*(L)$ or $\hat{\omega}_0(L)$ is a consistent estimator of $\omega_0^*(L)$. Here, $\hat{\omega}_0(L)$ may be obtained via a standard (linear) regression model for $E(Y|A, L)$; $\hat{\pi}(L)$ is typically obtained via a standard logistic regression model. The resulting (doubly-robust) G-estimator can be calculated as

$$\hat{\tau}_G(\phi) = \frac{\sum_{i=1}^{n} \{A_i - \hat{\pi}(L_i)\} \{Y_i - \phi\hat{\omega}_0(L_i)\}}{\sum_{i=1}^{n} \{1 - \hat{\pi}(L_i)\} A_i}.$$

The (doubly robust) inverse probability weighted (IPW) estimator ($\hat{\tau}_{IPW}(\phi)$) (Robins et al., 2000) is obtained as

$$\sum_{i=1}^{n} \frac{A_i}{\hat{\pi}(L_i)} \{Y_i - \phi\hat{\omega}_1(L_i)\} - \frac{1 - A_i}{1 - \hat{\pi}(L_i)} \{Y_i - \phi\hat{\omega}_0(L_i)\} + \phi \{\hat{\omega}_1(L_i) - \hat{\omega}_0(L_i)\},$$

where $\hat{\omega}_1(L)$ and $\hat{\omega}_0(L)$ are estimates of $E(Y|A = 1, L)$ and $E(Y|A = 0, L)$, respectively, based on possibly misspecified models. Again, ϕ is a user-specified constant. If set to 0, it yields the so-called IPW-estimator which is a CAN estimator of τ^* if $\hat{\pi}(L)$ is a consistent estimator of $\pi^*(L)$ for all L. If set to 1, it yields the so-called doubly-robust IPW-estimator which is a CAN estimator of τ^* if either $\hat{\pi}(L)$ is a consistent estimator of $\pi^*(L)$ for each L or $\hat{\omega}_j(L), j = 0, 1$ is

a consistent estimator of $E(Y|A=j,L)$ for each L (Leon et al., 2003).

Over the past decade, much attention has been given to the development of doubly robust estimation procedures (Robins and Rotnitzky, 2001). Facing the truth that in practice 'all' models are misspecified, the practical benefit of such doubly robust procedures has been questioned and concerns have been raised that such procedures may be very sensitive to misspecification affecting both nuisance working models (Kang and Schafer, 2008). We therefore evaluate the asymptotic bias (i.e., mean difference between the estimator and estimand) of the suggested G-estimators and IPW estimators under misspecification occurring in all nuisance working models. Upon using that the asymptotic bias of a root-n (asymptotically linear) estimator of τ^* with estimating function $U(\tau)$ equals $E\left\{\partial U(\tau^*)/\partial\tau\right\}^{-1}E\left\{U(\tau^*)\right\}$, we obtain asymptotic biases of

$$\frac{E\left[\left\{\pi^*(L)-\pi(L)\right\}\left\{\omega_0^*(L)-\phi\omega_0(L)\right\}\right]}{E\left[\left\{1-\pi(L)\right\}\pi^*(L)\right]} \tag{4.3}$$

for the G-estimator, and

$$E\left[\left\{\frac{\pi^*(L)}{\pi(L)}-1\right\}\left\{\omega_0^*(L)+\tau^*-\phi\omega_1(L)\right\}-\left\{\frac{1-\pi^*(L)}{1-\pi(L)}-1\right\}\left\{\omega_0^*(L)-\phi\omega_0(L)\right\}\right], \tag{4.4}$$

for the IPW-estimator. Here, $\pi(L),\omega_1(L)$ and $\omega_0(L)$ are the probability limits of $\hat{\pi}(L),\hat{\omega}_1(L)$ and $\hat{\omega}_0(L)$, respectively.

We will first focus on the estimators that set ϕ equal to zero. These are consistent estimators of τ^* under correct specification of the propensity score, but not necessarily otherwise. It is then seen that any degree of model misspecification in the propensity score of magnitude $\delta(L)=\pi^*(L)-\pi(L)$ at a given L, yields a contribution to the bias of the G-estimator of magnitude

$$\frac{\delta(L)\omega_0^*(L)}{E\left[\left\{1-\pi(L)\right\}\pi^*(L)\right]} \tag{4.5}$$

and to the bias of the IPW-estimator of magnitude

$$\delta(L)\left[\frac{\omega_0^*(L)}{\{1-\pi(L)\}\,\pi(L)}+\frac{\tau^*}{\pi(L)}\right]. \tag{4.6}$$

In the absence of an exposure effect (i.e. $\tau^*=0$), the bias contribution of the IPW-estimator is thus

$$\frac{E\left[\{1-\pi(L)\}\,\pi^*(L)\right]}{\{1-\pi(L)\}\,\pi(L)}$$

times that of the G-estimator. This ratio can be substantial within L-regions corresponding to propensity score values close to 0 or 1. Considering that such regions are typically located in the tails of the data distribution where model misspecification is more likely, we conclude that the IPW-estimator will generally be much more vulnerable than the G-estimator to misspecification of the propensity score.

Interestingly, the G-estimator can be consistent under misspecification of the propensity score model. This would happen for instance if the propensity score model were of the form $\pi(L)=\text{expit}(\alpha^{*'}L)$, α^* were estimated using a maximum likelihood procedure and $\omega_0^*(L)$ happened to be linear in L. In that case, the fitted propensity score would satisfy $0=E\left[\{A-\pi(L)\}\,\omega_0^*(L)\right]=E\left\{\delta(L)\omega_0^*(L)\right\}$, thus giving the estimating functions for the G-estimator, and in particular the bias term (4.5), mean zero. This is not the case for the IPW estimator when the propensity score is fitted using maximum likelihood inference as it follows from (4.6) that its bias due to propensity score misspecification depends on the magnitude of the exposure effect. Any cancelation of the bias of the IPW estimator under propensity score misspecification must thus be accidental in the sense of occurring only at one specific exposure effect size τ^*. Following a Bayesian argument (with an absolutely continuous prior density on τ), such cancelation occurs with zero probability. In view of this, it can be desirable to estimate the propensity score in such a way that consistency of the IPW-estimator is attained within a larger class of data-generating distributions than those that correspond to a correctly specified propensity score model. In particular,

we recommend calculating the IPW-estimator ($\hat{\tau}_{IPW}(\phi)$) as

$$\sum_{i=1}^{n} \frac{A_i}{\hat{\pi}_1(L_i)} \{Y_i - \phi\hat{\omega}_1(L_i)\} - \frac{1-A_i}{1-\hat{\pi}_0(L_i)} \{Y_i - \phi\hat{\omega}_0(L_i)\} + \phi \{\hat{\omega}_1(L_i) - \hat{\omega}_0(L_i)\},$$

where ϕ is as before, $\hat{\pi}_1(L_i)$ is a consistent estimator of $\pi(L_i)$ obtained by solving an estimating equation of the form

$$0 = \sum_{i=1}^{n} \left(\frac{A_i}{\pi(L_i)} - 1 \right) \varphi(L_i), \tag{4.7}$$

and $\hat{\pi}_0(L_i)$ is a consistent estimator of $\pi(L_i)$ obtained by solving an estimating equation of the form

$$0 = \sum_{i=1}^{n} \left(\frac{1-A_i}{1-\pi(L_i)} - 1 \right) \varphi(L_i), \tag{4.8}$$

where $\varphi(L_i)$ is an arbitrary index function of the dimension of α^*. Note that we use the same propensity score model, but different consistent estimators for the probability of exposure versus no exposure. Cancelation of the asymptotic bias may now occur when $\varphi(L)$ includes the constant 1 and $\omega_0^*(L)$ happens to be a linear combination of the components in the vector $\varphi(L)$. This can be seen from (4.4) with $\phi = 0$ upon noting that (4.7) and (4.8) then imply

$$E\left[\{\pi^*(L)/\pi(L) - 1\}\omega_0^*(L)\right] = E\left[\{(1-\pi^*(L))/(1-\pi(L)) - 1\}\omega_0^*(L)\right] = 0$$

. This would happen for instance if $\varphi(L) = (1, L)'$ and $\omega^*(L)$ happened to be linear in L. Note also that when $\varphi(L)$ includes the constant 1, then the asymptotic bias of the IPW-estimator is no longer dependent upon τ^* because it follows from equation (4.7) that $E\{\pi^*(L)/\pi(L)\}$ equals 1 in that case. In addition, the fitted propensity scores $\hat{\pi}_1(L_i)$ ($\hat{\pi}_0(L_i)$) are then such that the sum of the weights $1/\hat{\pi}_1(L_i)$ ($1/\{1-\hat{\pi}_0(L_i)\}$) in the exposed (unexposed) subjects equals the total sample size. We will therefore refer to estimation of the propensity scores following (4.7) and (4.8) as stabilized estimation. With a different attainment goal in mind, namely improving the stability of inverse weighting procedures,Cao et al. (2009) make a related, although different

proposal in a missing data context.

We will now focus on doubly-robust estimators obtained by setting $\phi = 1$. It is easily seen from both bias expressions (4.3) and (4.4) that any degree of model misspecification in the propensity score of magnitude $\delta(L) = \pi^*(L) - \pi(L)$ at a given L, and in the outcome regression models of magnitudes $\Delta_1(L) = \omega_0^*(L) + \tau^* - \omega_1(L)$ and $\Delta_0(L) = \omega_0^*(L) - \omega_0(L)$ yields a contribution to the bias of the doubly-robust G-estimator of magnitude

$$\frac{\delta(L)\Delta_0(L)}{E\left[\{1-\pi(L)\}\,\pi^*(L)\right]} \tag{4.9}$$

and to the bias of the doubly-robust IPW-estimator of magnitude

$$\delta(L)\left[\frac{\Delta_1(L)}{\pi(L)} - \frac{\Delta_0(L)}{1-\pi(L)}\right]. \tag{4.10}$$

It is immediate from these expressions that the doubly-robust G- and IPW-estimator have mean zero under misspecification of one, but not both nuisance working models. These estimators will typically also have smaller bias under propensity score misspecification than the previously considered G-estimator and IPW-estimator because any misspecification of magnitude $\delta(L)$ now gets inflated only proportional to the degree of misspecification in the outcome regression model.

Interestingly, the doubly-robust G-estimator not only is consistent under the union model which correctly specifies either the propensity score or the outcome regression, but also under certain data-generating mechanisms corresponding to misspecification affecting both nuisance working models. This would occur, for instance, if the misspecified propensity score model were of the form $\pi(L) = \text{expit}(\alpha_0 + \alpha_1 L + \alpha_2 L^2)$ and fitted using maximum likelihood inference, the fitted outcome regression model were of the form $\omega(L) = \gamma_0 + \gamma_1 L$ and $\omega^*(L)$ happened to be linear in L and L^2. Indeed, in that case the fitted propensity score model would satisfy $E\left[\{A-\pi(L)\}\,\{\omega_0^*(L) - \omega_0(L)\}\right] = E\left\{\delta(L)\Delta_0(L)\right\} = 0$. The doubly-robust IPW-estimator

with propensity scores fitted through maximum likelihood inference does not satisfy a similar property. In addition, it follows from (4.10) that misspecification in the regression model for $E(Y|A = 1, L)$ (or $E(Y|A = 0, L)$) can get dramatically inflated in L-regions where data on exposed subjects (on unexposed subjects) are relatively scarce. These are regions where model misspecification is also most likely, suggesting that doubly robust IPW-estimators may in fact exacerbate the extrapolation problem in view of which propensity-score adjusted estimators were designed. As a way of improving the performance of doubly robust estimators in the presence of influential weights, Robins et al. (2008) proposed fitting the outcome regression models $\omega_1(L)$ and $\omega_0(L)$, respectively, via standard weighted regression in the exposed and unexposed subjects, with weights $1/\pi(L)$ and $1/\{1 - \pi(L)\}$, respectively:

$$0 = \sum_{i=1}^{n} \frac{A_i}{\hat{\pi}(L_i)} \{Y_i - \hat{\omega}_1(L_i)\} \varphi_1(L_i) \tag{4.11}$$

$$0 = \sum_{i=1}^{n} \left\{ \frac{1 - A_i}{1 - \hat{\pi}(L_i)} \right\} \{Y_i - \hat{\omega}_0(L_i)\} \varphi_0(L_i), \tag{4.12}$$

where $\varphi_1(L_i)$ and $\varphi_0(L_i)$ are arbitrary vector functions of the dimension of the unknown parameters indexing $\omega_1(L_i)$ and $\omega_0(L_i)$, and including the constant 1. They refer to the resulting doubly robust estimator of the exposure effect as a regression doubly robust estimator. The advantage of this is clear from the fact that the above equations imply that

$$E\{(1 + \delta(L)/\pi_1(L))\Delta_1(L)\} = 0$$

and

$$E\{(1 - \delta(L)/(1 - \pi_0(L)))\Delta_0(L)\} = 0$$

so that the asymptotic bias of the doubly-robust IPW-estimator becomes

$$\delta(L) \left[\frac{\Delta_1(L)}{\pi(L)} - \frac{\Delta_0(L)}{1 - \pi(L)} \right] = -E\{\Delta_1(L) - \Delta_0(L)\}.$$

Bias due to model misspecification in the tails of the data distribution is thereby no longer inflated. Further robustness against model misspecification is attained by fitting the propensity score through equations (4.7) and (4.8), for then bias due to model misspecification cancels whenever $\Delta_1(L)$ and $\Delta_0(L)$ happen to be linear combinations of the components of $\varphi(L)$. This would occur, for instance, if the misspecified propensity score model were of the form $\pi(L) = \text{expit}(\alpha_0 + \alpha_1 L + \alpha_2 L^2)$, $\varphi(L) = (1, L, L^2)$, the fitted outcome regression model were of the form $\omega(L) = \gamma_0 + \gamma_1 L$ and $\omega_0^*(L)$ happened to be linear in L and L^2.

4.3.3 Simulation study

In this section, we illustrate the impact of global misspecification of the nuisance working models in G-estimators and IPW-estimators through a small simulation study. In each of 5000 simulation runs, a data set of 500 independent samples was generated with L a standard normal variate. In the first experiment, $Y = -2 + A + 2L + N(0,1)$ and $\pi^*(L) = \text{expit}(-3 + L)$. In the next 4 experiments, $Y = -2 + A + 2L - L^2 + N(0,1)$, with $\pi^*(L) = \text{expit}(-4 + 1.5\sqrt{|L|} + 0.75L + 0.5|L|^{1.5})$ in the second experiment, $\pi^*(L) = \text{expit}(-2 + 2\sin(2L))$ in the third experiment, and $\pi^*(L) = \text{expit}(-0.5 + \sin(2L) - 0.5\cos(3L) - 0.25L^2)$ in the fourth and fifth experiment. In all experiments, linear outcome working models were used. Second and third order logistic propensity score working models were used in the first three and last two experiments, respectively. Table 4.3.3 shows the bias and empirical standard deviation of the ordinary least squares estimates with (OLS-A) and without (OLS-U) adjustment for L, the G-estimator (G), the IPW-estimator with (IPW-S) and without (IPW) stabilized estimation of the propensity score, the regression doubly robust IPW estimator with (RDR-S) and without (RDR) stabilized estimation of the propensity score and the doubly robust IPW estimator with maximum likelihood estimation of the outcome working model and with (DR-S) and without (DR) stabilized estimation of the propensity score. The results demonstrate that in the absence of model misspecification (i.e. simulation experiment 1), stabilized estimation of the propensity score improves the finite-sample bias of the IPW estimator and yields a minor efficieny gain,

although an efficiency loss for the doubly robust estimators. In the presence of model misspecification, major improvements in both the bias and precision of (doubly robust) IPW estimators are observed. In particular, for the considered data-generating mechanisms, no bias was observed despite all working models being misspecified. The fourth and fifth experiment used the same data generating models, but $\varphi(L) = (1, L, |L|^{1.5}, L^2)$ in (4.7) and (4.8) in the fourth experiment and $\varphi(L) = (1, L, |L|^2, L^3)$ in the fifth experiment.

	Exp 1		Exp 2		Exp 3		Exp 4		Exp 5	
Estimator	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD
OLS-U	1.856	0.39	0.503	0.37	0.821	0.34	0.974	0.23	0.974	0.23
OLS-A	-0.002	0.18	-1.685	0.32	0.000	0.22	0.308	0.16	0.308	0.16
G	-0.009	0.2	0.002	0.16	-0.154	0.23	-0.063	0.11	-0.063	0.11
IPW	0.151	0.45	-0.285	0.32	-1.716	1.86	-0.123	0.82	-0.123	0.82
RDR	-0.005	0.28	-0.126	0.24	-0.363	0.20	0.014	0.30	0.014	0.30
DR	-0.006	0.29	-0.115	0.86	-4.863	122.51	-1.879	43.72	-1.879	43.72
IPW-S	0.028	0.42	0.003	0.19	0.361	0.50	0.027	0.13	0.043	0.15
RDR-S	-0.001	0.37	-0.027	0.21	-0.019	0.19	0.023	0.12	-0.004	0.12
DR-S	-0.001	0.37	-0.027	0.21	-0.044	0.18	0.022	0.12	0.000	0.12

Table 4.2: Simulation results: empirical bias and standard deviation in 5 simulation experiments.

4.4 Discussion

Modern procedures for marginal causal effects (see e.g. Section 4.3.2) require working models for the outcome and/or exposure, but their complexity does not affect the interpretability of the final effect estimand. The desire to use parsimonious models is therefore not so much stimulated by the need for obtaining interpretable results, but rather by concerns of bias and inefficiency which may result from overfitting. Two caveats are in place, however. First, while the possibility of bias resulting from overfitting is well understood for conditional effects (cfr. the Neyman-Scott paradox), to the best of our knowledge, the extent to which it affects the estimation of marginal effects remains to be evaluated. Second, it has been documented that efficiency gains may be realized when a priori knowledge is available that given covariates are only associated with the exposure, but have no residual association with outcome (Hahn, 2004). However,

in practice, such a priori information is rarely, if ever, available. Without such information, data-driven decisions must be made to exclude covariates from the analysis and it is unclear under what conditions the additional uncertainty induced by these selection approaches still enables a meaningful efficiency gain.

Most strategies used by practitioners to select confounders are based on excluding potential confounders from the analysis when they are non-significantly associated with the outcome conditional on the exposure; some focus on associations with the exposure instead. Such strategies are sub-optimal for various reasons. First, since confounders are by definition jointly associated with exposure and outcome, the importance of a variable as a confounder must ideally be judged against criteria that involve both associations. Second, even when for a given variable both associations are assessed, their significance is not directly informative about the extent to which adjusting for this variable will reduce confounding bias and, ultimately, improve the quality of the exposure effect estimator. Third, even when a more rigorous confounder-selection process is adopted, it remains difficult to acknowledge the uncertainty resulting from the selection process into the final inference. By ignoring this, one risks to obtain under-covering confidence intervals.

We have attempted to shed light on these issues and proposed a focused confounder-selection strategy which aims at minimum mean squared error of the exposure effect estimator. This strategy is closely linked to one recommended in Brookhart and van der Laan (2006), but computationally more attractive by avoiding the use of cross-validation. Its application overcomes the aforementioned first two concerns. In particular, when applied to estimators that are consistent under correct specification of a propensity score model, we expect it will overcome the usual difficulties (Brookhart et al., 2006) in selecting confounders in the propensity score model as the selection is made in terms of an 'optimal' trade-off between bias and efficiency of the exposure effect estimate and thus will have a tendency to 'automatically' exclude covariates that are solely associated with the exposure and include covariates that are solely associated with the outcome.

For such estimators, as shown in Section 4.2.6, it also roughly overcomes the third concern in the sense of retaining confidence validity even when the confounder-selection process is ignored. In spite of these attractions of focused confounder-selection based on propensity-score adjusted estimators, several limitations remain and warrant further study. First, the calculation of the mean squared error relies on estimates obtained from a full model which involves all potential confounders. Simulation studies are needed to evaluate finite-sample performance when these estimates are inefficient or biased as a result of overfitting. Second, in small samples, the procedure may choose to exclude potentially important confounders in order to reduce mean squared error at the expense of a bias, whose magnitude is difficult to assess. Stability plots like Figure 4.2 may help detect whether this occurs; one may use them, for instance, to restrict the procedure to all submodels that do not generate a bias exceeding a scientifically meaningful magnitude.

Given the aforementioned caveats and limitations of variable-selection, we see much value in the idea of avoiding confounder-selection by using regularization techniques such as ridge regression instead. This idea has been much advocated by Greenland (2007, 2008). Further research is needed to evaluate these contrasting viewpoints in realistic settings involving unmeasured confounding, missing confounder data and large separation in the confounder distributions of exposed and unexposed subjects. Perhaps the ideal future lies in an approach whereby the nuisance parameters indexing the working models for the association between covariates on the one hand, and exposure and outcome on the other hand, are estimated as those values that minimize the mean squared error of the exposure effect estimator. Such approach would combine the benefits of focused confounder-selection and regularization approaches that do not involve selection, and might improve upon them in various ways. In comparison with confounder-selection approaches, it would further lower the mean squared error by not being restricted to specific submodels and, by avoiding repeated model fitting, might enable a more easy assessment of the overall uncertainty. In comparison with approaches that involve no selection, it would have the advantage of directly targeting minimal mean squared error of the exposure effect estimator. It

is unclear at present whether such approach is attainable.

Appendix

Assessment of M-bias and bias amplification

Consider the path diagram in Figure 4.1. Let Y^*, A^* and L^* denote standardized (Wright, 1934) variables corresponding to Y, A and L, respectively. Assume that $E(Y^*|A^*, L^*, U_2, U_3) = cA^*+bL^*+c_{2y}U_2+c_{3y}U_3$, then $E(Y^*|A^*, L^*) = cA^*+bL^*+c_{2y}E(U_2|A^*, L^*)+c_{3y}E(U_3|A^*, L^*)$. Let $E(U_2|A^*, L^*) = \alpha_2 L^* + \beta_2 A^*$, $E(L^*|U_1, U_2) = c_{1l}U_1 + c_{2l}U_2$ and $E(A^*|L^*, U_1, U_3) = c_{1a}U_1 + c_{3a}U_3 + aL^*$ Then, proceeding as in Pearl (Pearl, 2010), we have $E(U_2L^*) \equiv c_{2l} = \alpha_2 + \beta_2\rho_{al}$ and $E(U_2A^*) \equiv c_{2l}a = \alpha_2\rho_{al} + \beta_2$, where $\rho_{al} = a + c_{1a}c_{1l}$, from which

$$\alpha_2 = c_{2l}\frac{(1-\rho_{al}^2+c_{1a}c_{1l}\rho_{al})}{1-\rho_{al}^2}\ ,\ \beta_2 = -c_{2l}\frac{c_{1a}c_{1l}}{1-\rho_{al}^2}.$$

Likewise, $E(U_3|A^*, L^*) = -c_{3a}\rho_{al}/(1-\rho_{al}^2)L^* + c_{3a}/(1-\rho_{al}^2)A^*$. It follows that

$$\begin{aligned}E(Y^*|A^*, L^*) = &\left(b + c_{2y}c_{2l}\frac{(1-\rho_{al}^2+c_{1a}c_{1l}\rho_{al})}{1-\rho_{al}^2} - \frac{c_{3a}c_{3y}\rho_{al}}{1-\rho_{al}^2}\right)L^* \\ &+ \left(c - c_{2y}c_{2l}\frac{c_{1a}c_{1l}}{1-\rho_{al}^2} + \frac{c_{3y}c_{3a}}{1-\rho_{al}^2}\right)A^*,\end{aligned}$$

and $E(Y^*|A^*) = \{(b + c_{2y}c_{2l})\rho_{al} + c - c_{2y}c_{2l}c_{1a}c_{1l} + c_{3y}c_{3a}\}A^*$. The bias reported in the main text is the difference between the coefficient joining A^* in the above expressions, and the population causal effect c. It is easy to demonstrate that inverse weighting by $1/f(A^*|L^*)$ yields an exposure-outcome covariance equal to

$$\int YAf(Y|A, L)f(L)dYdAdL = \left(c - c_{2y}c_{2l}\frac{c_{1a}c_{1l}}{1-\rho_{al}^2} + \frac{c_{3y}c_{3a}}{1-\rho_{al}^2}\right).$$

In Figure 4.4, we develop a better understanding of the magnitude of these biases under the assumption that ρ_2, as defined in the main text, is at most ρ_{al}. We make this assumption to respect that, arguably, U_3 will have weaker correlations with exposure and outcome than L when the focus of the study is on assessing the effect of A on Y, as efforts have then been targeted

at collecting data on common causes of exposure and outcome. We make a similar assumption for ρ_1 to respect the fact that ρ_1 indirectly contributes to the magnitude of ρ_{al}. The solid line in Figure 4.4 displays the upper bound (4.1) in a setting where $\rho_1 = \rho_{al}/2$ and $\rho_2 = \rho_{al}/3$. It shows that the adjusted analysis will only be more biased than the unadjusted analysis when the correlation between A and L is extremely large and the correlation between Y and L is extremely small. We believe this is unlikely to occur in practice. The figure further suggests that, under the considered scenario, the impact of M-bias (see bottom line in Figure 4.4) is not much less sizeable than that of unmeasured confounding (see top line in Figure 4.4), although only of importance for exposure-confounder correlations exceeding 0.5.

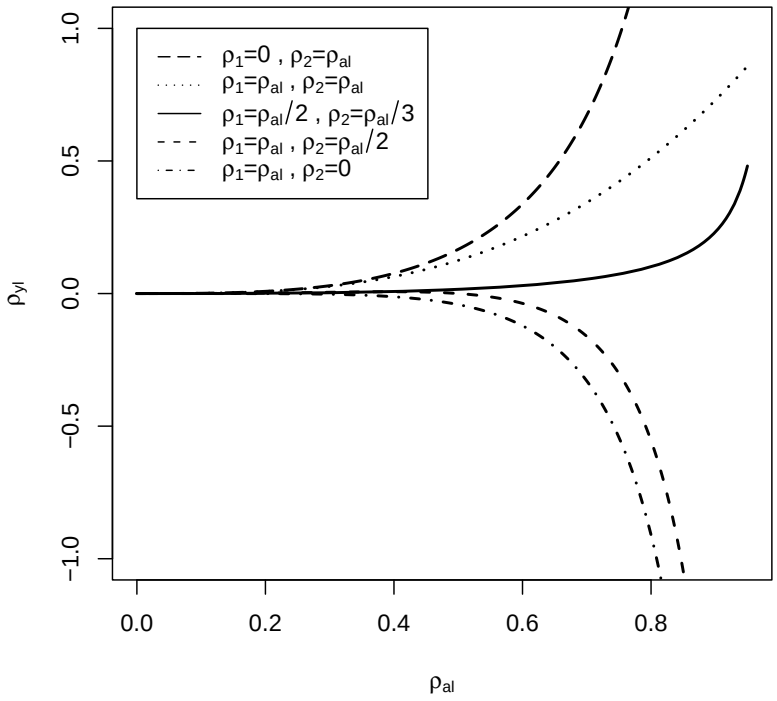

Figure 4.4: Values of ρ_{yl} below which the adjusted analysis has larger bias than the unadjusted analysis.

Implementation in standard software

In section 4.2.5 we evaluate the proposed confounder-selection procedure in an observational study investigating the effect of right heart catherization (RHC) on 180 day mortality in 5735

critically ill patients (Connors et al., 1996). In total, the *confounder space* consists of 61 covariates (L) or 62 *orbits* with orbit 1 equal to the narrow model (only intercept and exposure of interest) and orbit 62 equals the full model (including the 61 confounders together with the exposure of interest). Throughout the analysis we explore all orbits and select the j^{th} orbit (with $j-1$ confounders) for which the *quality* (in terms of mean squared error) of the exposure effect estimator is the best. We start by fitting a marginal model (orbit 1). From orbit 1 we then move to orbit 2 by selecting 1 covariate that minimizes the MSE of the effect of interest. We transfer *the best covariate* to orbit 3. In orbit 3 we choose a second covariate that minimizes the MSE of the effect of interest. Then, a stochastic search selects at random one covariate which is in the model and one which is not, and constructs a new model by interchanging both covariates. The procedure continues until we arrive at the last orbit in which a full model is fitted to the data.

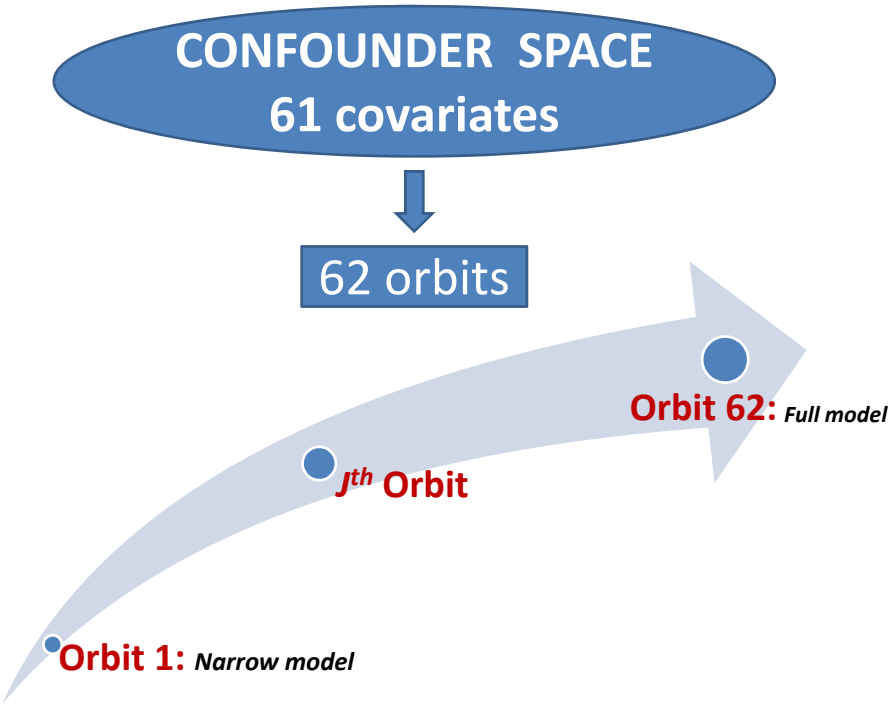

Figure 4.5: Exploration of the confounder space

In this section we provide the syntax in R for:

- The algorithm for the exploration of the orbits and the stochastic search (adapted from (Crainiceanu et al., 2008)).
- The calculation of the MSE using an asymptotic approximation which can be made under a local misspecification assumption.

Calculation of the MSE

A framework for this is developed in Hjort and Claeskens (Hjort and Claeskens, 2003) for parametric models and adapted to our specific setting. For the calculation of the MSE we wrote a user friendly function:

```
MSE=function(design.matrix,full.model,sub.model)
```

which gives an estimate of the marginal log odds ratio together with a SE and 95% CI. It also provides the MSE and FIC selection criterium. As an argument from the function one needs to provide a dataset with all confounders `design.matrix.nox` , the full model `full.model` and a model with a subset of confounders from which one wishes to know the MSE or FIC `sub.model`.

What happens inside this function is further explained in the remainder of this section. We consider the marginal log odds ratio as a focus parameter, which we define as

$$\tau^* = \log \frac{\mu_1(1-\mu_0)}{\mu_0(1-\mu_1)}$$

where $\mu_a = E\left[\text{expit}\left\{\omega(L;\gamma^*) + \beta_0^* + \beta_a^* a\right\}\right]$ for $a = 0, 1$. Denote furthermore

$$\hat{\mu}_a = n^{-1} \sum_{i=1}^{n} \left[\text{expit}\left\{\omega(L_i;\hat{\gamma}) + \hat{\beta}_0 + \hat{\beta}_a a\right\}\right]$$

for $a = 0, 1$.

In R, several steps are needed to calculate the the marginal log odds ratio. In a first step we construct an index `sub.ind` indicating which confounders are included in the submodel.

```
sub.ind=(names(design.matrix) %in%
      attr(summary(sub.model)$terms,"term.labels")[-1])
Xdat.sub=design.matrix[sub.ind]
```

In a next step we calculate μ_a for $a = 0, 1$ and obtain the marginal log odds ratio `MLOR` for a given submodel S.

```
Xdat_1.sub=cbind(1,1,Xdat.sub)
Xdat_0.sub=cbind(1,0,Xdat.sub)
p1.sub=expit(coef(sub.model)%*% t(Xdat_1.sub))
p0.sub=expit(coef(sub.model) %*% t(Xdat_0.sub))
mu1.sub=mean(p1.sub)
mu0.sub=mean(p0.sub)
MLOR=log((mu1.sub*(1-mu0.sub))/(mu0.sub*(1-mu1.sub)))
```

Assume, as in Claeskens and Hjort (2003), that the true data density $f(Y, A|L)$ is indexed by a parameter $\beta^* = (\beta_0^*, \beta_a^*)'$, which is shared between all models, and $\gamma^* + \delta/\sqrt{n}$, where the term $\delta/\sqrt{n}$ encodes local model misspecification (see Section 4.2.6) and γ^* is the vector of values to which the nuisance parameter γ is set in the narrow model (that is, typically $\gamma^* = 0$). Let further $\theta_S \equiv (\beta, \gamma_S)'$ and $\hat{\theta}_S \equiv (\hat{\beta}_S, \hat{\gamma}_S)'$. Then we have that for any submodel S, the corresponding estimator $\hat{\mu}_{Sa}$ of μ_a^* (which is defined like $\hat{\mu}_a$, but with $\hat{\gamma}_S$ and $\hat{\beta}_S$ replacing $\hat{\gamma}$

and $\hat{\beta}$, respectively) satisfies

$$\begin{aligned}
0 &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\text{expit}\left\{\omega(L_i;\hat{\gamma}_S,\gamma^*_{-S})+\hat{\beta}_{S0}+\hat{\beta}_{Sa}a\right\}-\sqrt{n}\hat{\mu}_{Sa}\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\text{expit}\left\{\omega(L_i;\gamma^*+\delta/\sqrt{n})+\beta^*_0+\beta^*_a a\right\}-\sqrt{n}\mu^*_a\\
&\quad + E\left[\frac{\partial}{\partial\theta_S}\text{expit}\left\{\omega(L_i;\gamma^*+\delta/\sqrt{n})+\beta^*_0+\beta^*_a a\right\}\right]\sqrt{n}\left(\hat{\theta}_S-\theta^*_S\right)\\
&\quad - E\left[\frac{\partial}{\partial\gamma}\text{expit}\left\{\omega(L_i;\gamma^*+\delta/\sqrt{n})+\beta^*_0+\beta^*_a a\right\}\right]\delta-\sqrt{n}(\hat{\mu}_{Sa}-\mu^*_a)+o_p(1)
\end{aligned}$$

from which

$$\begin{aligned}
\sqrt{n}(\hat{\mu}_{Sa}-\mu^*_a) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\text{expit}\left\{\omega(L_i;\gamma^*+\delta/\sqrt{n})+\beta^*_0+\beta^*_a a\right\}-\mu^*_a\\
&\quad + E\left[\frac{\partial}{\partial\theta_S}\text{expit}\left\{\omega(L_i;\gamma^*+\delta/\sqrt{n})+\beta^*_0+\beta^*_a a\right\}\right]\sqrt{n}\left(\hat{\theta}_S-\theta^*_S\right)\\
&\quad - E\left[\frac{\partial}{\partial\gamma}\text{expit}\left\{\omega(L_i;\gamma^*+\delta/\sqrt{n})+\beta^*_0+\beta^*_a a\right\}\right]\delta+o_p(1).
\end{aligned}$$

It follows from the Delta method that the influence function (Tsiatis, 2006) for $\hat{\tau}_S$ is $D_\mu + d_\beta\sqrt{n}(\hat{\beta}_S-\beta^*)+d_{\gamma_S}\sqrt{n}(\hat{\gamma}_S-\gamma^*_S)-d_\gamma\delta$, where

$$\begin{aligned}
d_\gamma &= \frac{1}{\mu^*_1(1-\mu^*_1)}E\left[\frac{\partial}{\partial\gamma}\text{expit}\left\{\omega(L_i;\gamma^*+\delta/\sqrt{n})+\beta^*_0+\beta^*_a\right\}\right]\\
&\quad - \frac{1}{\mu^*_0(1-\mu^*_0)}E\left[\frac{\partial}{\partial\gamma}\text{expit}\left\{\omega(L_i;\gamma^*+\delta/\sqrt{n})+\beta^*_0\right\}\right]\\
D_\mu &= \frac{1}{\mu^*_1(1-\mu^*_1)}\left[\text{expit}\left\{\omega(L_i;\omega(L_i;\gamma^*+\delta/\sqrt{n})+\beta^*_0+\beta^*_a\right\}-\mu^*_1\right]\\
&\quad - \frac{1}{\mu^*_0(1-\mu^*_0)}\left[\text{expit}\left\{\omega(L_i;\gamma^*+\delta/\sqrt{n})+\beta^*_0\right\}-\mu^*_0\right],
\end{aligned}$$

and where d_β and d_{γ_S} are defined like d_γ, but with derivatives taken w.r.t. β and γ_S, respectively, rather than γ.

```
Xdat_1.full=cbind(1,1,design.matrix)
```

```
Xdat_0.full=cbind(1,0,design.matrix)
p1=expit(coef(full.model) %*% t(Xdat_1.full))
p0=expit(coef(full.model) %*% t(Xdat_0.full))
mu1=mean(p1)
mu0=mean(p0)

d_beta=(1/(mu1*(1-mu1)))*mean(p1*(1-p1)*1)
d_int=(1/(mu1*(1-mu1)))*mean(p1*(1-p1)*1)-
      (1/(mu0*(1-mu0)))*mean(p0*(1-p0)*1)
d_beta=c(d_int,d_beta)

if(sum(sub.ind)==0){
d_theta=d_beta}
if (sum(sub.ind) !=0){
tmp1.sub=apply(t(as.matrix(as.vector(p1.sub*(1-p1.sub))))*
         (Xdat.sub),2,mean)
tmp0.sub=apply(p0.sub*(1-p0.sub)*Xdat.sub,2,mean)
d_gammas=(1/(mu1.sub*(1-mu1.sub)))*tmp1.sub-(1/(mu0.sub*
          (1-mu0.sub)))*tmp0.sub
d_theta=c(d_beta,d_gammas)
}

tmp1=apply(p1*(1-p1)*design.matrix,2,mean)
tmp0=apply(p0*(1-p0)*design.matrix,2,mean)
d_gamma=(1/(mu1*(1-mu1)))*tmp1-(1/(mu0*(1-mu0)))*tmp0

D_mu=as.numeric(1/(mu1*(1-mu1))*(p1-mu1)-1/(mu0*(1-mu0))*(p0-mu0))
```

```
VAR_DMU=var(D_mu)
```

With the information obtained above we can calculate the variance of the marginal log odds ratio

```
COV=summary(sub.model)$cov.unscaled
SE=sqrt(VAR_DMU/n + d_theta%*%COV%*%t(t(d_theta)))
CI_95=c(MLOR-1.96*SE,MLOR+1.96*SE)
```

Using Lemmas 3.2 and 3.3 in Hjort and Claeskens (2003), it can be shown that $\sqrt{n}(\hat{\tau}_S - \tau^*)$ converges in distribution to $\Lambda_0 + \omega'(\delta - G_S D)$, where

$$\begin{aligned}
\Lambda_0 &= d_\beta J_{00}^{-1} M' + D_\mu \\
D &= \delta + Q(N' - J_{10} J_{00}^{-1} M') \\
Q &= \left(J_{11} - J_{10} J_{00}^{-1} J_{01}\right)^{-1} \\
\omega &= J_{10} J_{00}^{-1} d_\beta - d_\gamma \\
G_S &= \pi_S \left\{\pi_S' Q^{-1} \pi_S\right\}^{-1} \pi_S Q^{-1}
\end{aligned}$$

with π_S the projection matrix for submodel S (i.e., a matrix of zeros with as many rows and columns as the dimensions of γ_S and γ, respectively, and with a 1 on each row in the column representing the corresponding component of γ_S), (M', N') following a mean zero normal distribution with covariance matrix

$$J = \begin{pmatrix} J_{00} & J_{01} \\ J_{10} & J_{11} \end{pmatrix},$$

which is 1 over n times the inverse of the asymptotic covariance matrix of $\hat{\theta} = (\hat{\beta}, \hat{\gamma})'$. These expressions rely on $\hat{\theta}$ being a maximum likelihood estimator.

```
COV_full=summary(full.model)$cov.unscaled
```

```
J=solve(COV_full)/n

J00=as.matrix(J[c(1,2),c(1,2)])
J10=as.matrix(J[,c(1:2)][-c(1:2),])
J01=as.matrix(J[c(1:2),][,-c(1:2)])
J11=as.matrix(J[-c(1:2),-c(1:2)])

var_LAMBDA0= d_beta%*%solve(J00)%*%d_beta + VAR_DMU
omega=t(J10%*%solve(J00)%*%as.matrix(d_beta))-d_gamma
Q=solve(J11-(J10%*%solve(J00)%*%J01))
```

The projection matrix π_S the submodel S is calculated as follows:

```
PM=diag(dim(design.matrix)[2])
diag(PM)=sub.ind*1
PM=as.matrix(PM[apply(PM,1,sum)==1,])
```

and is used in the calculation of G_S

```
if(sum(sub.ind)==1){
GS=PM%*%solve(t(PM)%*%(J11-(J10%*%solve(J00)%*%J01))%*%PM)%*%
        t(PM)%*%(J11-(J10%*%solve(J00)%*%J01))
}
if(sum(sub.ind)==0){
GS=matrix(nrow=length(sub.ind),ncol=length(sub.ind),0)
}
if (sum(sub.ind)>1){
GS=t(PM)%*%solve(PM%*%(J11-(J10%*%solve(J00)%*%J01))%*%t(PM))%*%
```

```
        PM%*%(J11-(J10%*%solve(J00)%*%J01))
}
```

Further, Λ_0 can be shown to be uncorrelated with D because M is independent of D by Lemma 3.3 (Hjort and Claeskens, 2003), and because D_μ is uncorrelated with D by the fact that (a) $Q(N' - J_{10}J_{00}^{-1}M')$ is the asymptotic distribution of $\sqrt{n}(\hat{\gamma} - \gamma^* - \delta/\sqrt{n})$; and that (b) the influence functions of $\hat{\gamma}$ are uncorrelated with D_μ by the fact that the former have mean zero conditional on L, whilst the latter are functions of L. It now follows that $\sqrt{n}(\hat{\tau}_S - \tau^*)$ has limiting mean squared error given by $\text{Var}(\Lambda_0) + \omega' G_S Q G_S' \omega + \omega'(I - G_S)\delta\delta'(I - G_S)'\omega$. Upon substituting (Claeskens and Hjort, 2008) $\delta\delta'$ with $\max(0, D_n D_n' - \hat{Q})$, where $D_n = \sqrt{n}(\hat{\gamma} - \gamma^*)$, we obtain

$$\text{Var}(\Lambda_0) - \omega' Q \omega + 2\omega' G_S Q G_S' \omega + \omega'(I - G_S) D_n D_n' (I - G_S)'\omega.$$

Because the first two terms are common to all models, we employ the remaining terms as a Focused Information Criterion (Claeskens and Hjort, 2003, 2008), upon substituting all population values with consistent estimates.

So finally everything is combined and the MSE and FIC are calculated as follows:

```
Dn=sqrt(n)*parms[-c(1,2)]
delta=Dn%*%t(Dn)-Q
delta1=Dn%*%t(Dn)

VAR= (var_LAMBDA0 + omega%*%GS%*%Q%*%t(GS)%*%t(omega))/n
BIAS2=((omega%*%(diag(dim(GS)[1])-GS)%*%delta1%*%
        t((diag(dim(GS)[1])-GS))%*%t(omega)))/n

mse=VAR+BIAS2)
```

```
FIC=2*(omega%*%GS%*%Q%*%t(GS)%*%t(omega)) + ((omega)%*%
(diag(dim(GS)[1])-GS)%*%delta1%*%t((diag(dim(GS)[1])-GS))%*%t(omega))
```

Orbit selection algorithm

We adapted the algorithm of Crainiceanu et al. (2008) and wrote a function to explore all the orbits of a model space generated by an outcome and a list of covariates. The orbit can be seen as the collection of models with the same number of covariates. The function starts from orbit 0 (no additional covariates) and iterates through orbits until the last orbit (containing all the covariates). Within each orbit the function searches for model which reduces MSE/FIC of the targeted effect. This is done in two phases. Phase one (the so called Project and Pursuit) is a greedy search phase. Using the covariates identified on the previous orbit the model is looking for the one additional covariate that reduces the MSE/FIC of the targeted effect. Phase 2 is a fuzzy search phase. It starts with the best model identified in the greedy phase. At each step it uses an MCMC update where any variable can enter or leave the model and again searches for the model that reduces the MSE/FIC of the targeted effect.

```
Orbit <- function(n.iterations,family.name, always.in.var,
all.xs, dependent.variable,data,full.model)
```

`n.iterations` indicates the number of iterations in the fuzzy phase within each orbit. The outcome regression family is given by `family.name`. All regression families accepted by the `glm()` function are acceptable. `always.in.var` gives vector of labels of variables in the original data set for those variables that are always in the model. It can be NULL. The labels of the potential confounders which can be selected in each orbit and define the entire model space that will be explored are given in `all.xs`and can be in or out of the model. The dependent variable is given by `dependent.variable` for the original data set `data`. Finally in the calculation of the MSE/FIC a full model is used and defined in `full.model`. The function to calculate the MSE/FIC is embedded into the `orbit function`.

The full R-code for the functions can be obtained upon request from `maarten.bekaert@ugent.be` or `stijn.vansteelandt@ugent.be`

Model uncertainty

Let $U(\tau, \alpha)$ be the estimating function for τ and $S_S(\alpha) = \partial \log f(A|L;\alpha)/\partial \alpha_S$, where α_S is the subvector of α which is free under model S and α_{-S} is the remaining part. Let $\hat{\tau}_S$ denote the estimator of τ^* as obtained under model S, and $\sum_{S\in\mathcal{A}} c(S|D_n)\hat{\tau}_S$ denote the estimator of τ^* obtained under model selection, where the weight $c(S|D_n)$ assigns 1 to the selected model and 0 to all other models and where $\mathcal{A}$ denotes the model space. Under the local misspecification assumption, we have that (Claeskens and Hjort, 2003)

$$\sqrt{n}\left\{\sum_{S\in\mathcal{A}} c(S|D_n)\hat{\tau}_S - \tau^*\right\} \xrightarrow{d} \sum_{S\in\mathcal{A}} c(S|D)\Lambda_S,$$

where Λ_S is the limit distribution of $\sqrt{n}\,(\hat{\tau}_S - \tau^*)$. Under this assumption, a Taylor series expansion shows that

$$\begin{aligned}
0 &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i(\hat{\tau}_S, \hat{\alpha}_S, \alpha^*_{-S}) \\
&= \left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i(\tau^*, \alpha^*_{Sn}, \alpha^*_{-Sn})\right\} + E\left(\frac{\partial}{\partial\tau}U_i(\tau^*, \alpha^*_{Sn}, \alpha^*_{-Sn})\right)\sqrt{n}\,(\hat{\tau}_S - \tau^*) \\
&\quad + E\left(\frac{\partial}{\partial\alpha_S}U_i(\tau^*, \alpha^*_{Sn}, \alpha^*_{-Sn})\right)\sqrt{n}\,(\hat{\alpha}_S - \alpha^*_{Sn}) - E\left(\frac{\partial}{\partial\alpha_{-S}}U_i(\tau^*, \alpha^*_{Sn}, \alpha^*_{-Sn})\right)\delta_{-S} + o_p(1),
\end{aligned}$$

where δ_{-S} is the subvector of δ corresponding to α_{-S} and $\alpha_{Sn} = \alpha_S + \delta_S/\sqrt{n}$ and $\alpha_{-Sn} = \alpha_{-S} + \delta_{-S}/\sqrt{n}$. Likewise, we have that

$$\begin{aligned} 0 &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} S_{Si}(\hat{\alpha}_S, \alpha^*_{-S}) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} S_{Si}(\alpha^*_{Sn}, \alpha^*_{-Sn}) + E\left(\frac{\partial}{\partial \alpha_S} S_{Si}(\alpha^*_{Sn}, \alpha^*_{-Sn})\right) \sqrt{n}\,(\hat{\alpha}_S - \alpha^*_{Sn}) \\ &\quad - E\left(\frac{\partial}{\partial \alpha_{-S}} S_{Si}(\tau^*, \alpha^*_{Sn}, \alpha^*_{-Sn})\right) \delta_{-S} + o_p(1), \end{aligned}$$

from which

$$\begin{aligned} \sqrt{n}\,(\hat{\tau}_S - \tau^*) \;=\; & E\left(\frac{\partial}{\partial \tau} U_i(\tau^*, \alpha^*_{Sn}, \alpha^*_{-Sn})\right)^{-1} \left[-\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ U_i(\tau^*, \alpha^*_{Sn}, \alpha^*_{-Sn}) \right.\right. \\ & -E\left(\frac{\partial}{\partial \alpha_S} U_i(\tau^*, \alpha^*_{Sn}, \alpha^*_{-Sn})\right) E\left(\frac{\partial}{\partial \alpha_S} S_{Si}(\alpha^*_{Sn}, \alpha^*_{-Sn})\right)^{-1} S_{Si}(\alpha^*_{Sn}, \alpha^*_{-Sn}) \Big\} \\ & +\delta_{-S} \left\{ E\left(\frac{\partial}{\partial \alpha_{-S}} U_i(\tau^*, \alpha^*_{Sn}, \alpha^*_{-Sn})\right) - E\left(\frac{\partial}{\partial \alpha_S} U_i(\tau^*, \alpha^*_{Sn}, \alpha^*_{-Sn})\right) \right. \\ & \left.\left. \times E\left(\frac{\partial}{\partial \alpha_S} S_{Si}(\alpha^*_{Sn}, \alpha^*_{-Sn})\right)^{-1} E\left(\frac{\partial}{\partial \alpha_{-S}} S_{Si}(\alpha^*_{Sn}, \alpha^*_{-Sn})\right) \right\} \right] + o_p(1). \end{aligned}$$

Further, $E\left\{U_i(\tau^*, \alpha^*_{Sn}, \alpha^*_{-Sn})\right\} = 0$ implies that $E\left\{\partial U_i(\tau^*, \alpha^*_{Sn}, \alpha^*_{-Sn})/\partial \alpha_S\right\}$ equals $-E\left\{U_i(\tau^*, \alpha^*_{Sn}, \alpha^*_{-Sn}) S_{Si}(\alpha^*_{Sn}, \alpha^*_{-Sn})\right\}$ and likewise for $S_{Si}(\alpha^*_{Sn}, \alpha^*_{-Sn})$. We thus find that

$$\begin{aligned} \sqrt{n}\,(\hat{\tau}_S - \tau^*) \;=\; & E\left(\frac{\partial}{\partial \tau} U_i(\tau^*, \alpha^*_{Sn}, \alpha^*_{-Sn})\right)^{-1} \left[-\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ U_i(\tau^*, \alpha^*_{Sn}, \alpha^*_{-Sn}) \right.\right. \\ & -E\left\{U_i(\tau^*, \alpha^*_{Sn}, \alpha^*_{-Sn}) S_{Si}(\alpha^*_{Sn}, \alpha^*_{-Sn})\right\} E\left\{S^{\otimes 2}_{Si}(\alpha^*_{Sn}, \alpha^*_{-Sn})\right\}^{-1} S_{Si}(\alpha^*_{Sn}, \alpha^*_{-Sn}) \Big\} \\ & +\delta_{-S} \left\{ E\left(\frac{\partial}{\partial \alpha_{-S}} U_i(\tau^*, \alpha^*_{Sn}, \alpha^*_{-Sn})\right) - E\left(\frac{\partial}{\partial \alpha_S} U_i(\tau^*, \alpha^*_{Sn}, \alpha^*_{-Sn})\right) \right. \\ & \left.\left. \times E\left(\frac{\partial}{\partial \alpha_S} S_{Si}(\alpha^*_{Sn}, \alpha^*_{-Sn})\right)^{-1} E\left(\frac{\partial}{\partial \alpha_{-S}} S_{Si}(\alpha^*_{Sn}, \alpha^*_{-Sn})\right) \right\} \right] + o_p(1), \end{aligned}$$

where for an arbitrary matrix, $A^{\otimes 2} \equiv AA'$. It then follows that $\sqrt{n}\left\{\sum_{S\in\mathcal{A}} c(S|D_n)\hat{\tau}_S - \tau^*\right\}$ is

$$
\begin{aligned}
&E\left(\frac{\partial}{\partial\tau}U_i(\tau^*,\alpha^*_{Sn},\alpha^*_{-Sn})\right)^{-1}\left[-\sum_{S\in\mathcal{A}} c(S|D)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{U_i(\tau^*,\alpha^*_{Sn},\alpha^*_{-Sn})\right.\right.\\
&\left.-E\left\{U_i(\tau^*,\alpha^*_{Sn},\alpha^*_{-Sn})S_{Si}(\alpha^*_{Sn},\alpha^*_{-Sn})\right\}E\left\{S^{\otimes 2}_{Si}(\alpha^*_{Sn},\alpha^*_{-Sn})\right\}^{-1}S_{Si}(\alpha^*_{Sn},\alpha^*_{-Sn})\right\}\\
&+\sum_{S\in\mathcal{A}} c(S|D)\delta_{-S}\left\{E\left(\frac{\partial}{\partial\alpha_{-S}}U_i(\tau^*,\alpha^*_{Sn},\alpha^*_{-Sn})\right)-E\left(\frac{\partial}{\partial\alpha_S}U_i(\tau^*,\alpha^*_{Sn},\alpha^*_{-Sn})\right)\right.\\
&\left.\left.\times E\left(\frac{\partial}{\partial\alpha_S}S_{Si}(\alpha^*_{Sn},\alpha^*_{-Sn})\right)^{-1}E\left(\frac{\partial}{\partial\alpha_{-S}}S_{Si}(\alpha^*_{Sn},\alpha^*_{-Sn})\right)\right\}\right]+o_p(1).
\end{aligned}
$$

It now follows by the Cauchy-Schwarz inequality that an upper bound to the asymptotic variance of $\sqrt{n}\left\{\sum_{S\in\mathcal{A}} c(S|D_n)\hat{\tau}_S - \tau^*\right\}$ is the variance of

$$
E\left(\frac{\partial}{\partial\tau}U_i(\tau^*,\alpha^*_{Sn},\alpha^*_{-Sn})\right)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i(\tau^*,\alpha^*_{Sn},\alpha^*_{-Sn}).
$$

It does not immediately follow that standard confidence intervals based on this conservative variance estimate will themselves be conservative. This is because $\sum_{S\in\mathcal{A}} c(S|D_n)\hat{\tau}_S$ follows a mixture distribution with bias components converging at root-n rate to zero. Because misspecifications δ of the order 1 over root-n are not consistently estimable (Claeskens and Hjort, 2003), there is little room for further correcting this, unless for instance a doubly robust estimator with correctly specified nuisance outcome working model happens to be used, in which case uncertainty in the propensity score model does not affect inferences for τ^*.

5

Estimation of natural direct and indirect effects

The work in this chapter is submitted and currently under review. The general idea about the simultaneous estimation of natural direct and indirect effects using the simple IPW estimator was independently developed by Theis Lange and submitted as a *proof of concept paper*. In a follow up paper, from which this chapter is adapted, we propose additional estimators for the natural direct and indirect effect which can be simply implemented in standard software.

Summary

Mediation analysis is widely adopted to infer causal mechanism by disentangling indirect or

mediated effects of an exposure on an outcome through given intermediaries, from the remaining direct effect. Traditional approaches build on standard regression models for the outcome and mediator, but easily result in difficult-to-interpret or difficult-to-report results when some of these models involve non-linearities. In this article, we focus on a broad class of so-called natural effect models whose parameters encode both (natural) direct and indirect effects (mediated through a given intermediary). We propose flexible estimation strategies for the parameters indexing these models, that are easy to perform with standard statistical software: an inverse probability weighted estimator, a regression mean imputation estimator and a doubly robust estimator. We give a theoretical discussion of the properties of these estimators and assess their finite-sample performance through a simulation study and through the analysis of the WHO-LARES study on the association between residence in a damp and moldy dwelling and the risk of depression.

5.1 Introduction

Researchers in a variety of scientific fields, notably epidemiologists, have a longstanding interest in using empirical studies to disentangle causal pathways by which an exposure or treatment affects an outcome. Most popular is the use of mediation analysis (MacKinnon, 2008), whereby an exposure's effect is decomposed into an indirect effect mediated by a given intermediate variable and the remaining direct effect. Prevailing methods for mediation analysis have their roots in work by Baron and Kenny (1986), who focused on linear models for the outcome and mediator. Baron and Kenny (1986) proposed estimating the direct effect as the residual association between outcome and exposure after regression adjustment for the mediator(s). They further proposed estimating the indirect effect as the product of the exposure's effect on the mediator and the mediator's effect on the outcome, or equivalently, as the difference between the total effect and the direct effect, where the former refers to the (unadjusted) association between outcome and exposure, respectively. Both these approaches have subsequently been employed in a

variety of statistical models other than the linear model and are typically referred to as 'product of coefficient methods' and 'difference in coefficient methods' for indirect effects. Despite their popularity, a theoretical basis beyond the linear model is lacking. Not only may these methods fail to capture a measure of indirect effect, also the interpretation of the estimates that they return, is often unclear (see e.g. VanderWeele and Vansteelandt (2009); VanderWeele and Vansteelandt (2010)).

Robins and Greenland (1992) revolutionized mediation analysis by proposing a generic strategy to decompose a total effect into a so-called pure or natural direct and indirect effect, that is not tied to a particular statistical model. Under appropriate identification conditions, use of the so-called mediation formula (Pearl, 2001, 2011) then enables combining arbitrary statistical models for the outcome and mediator to obtain valid and well-understood measures of direct and indirect effect (see e.g. VanderWeele and Vansteelandt (2009); VanderWeele and Vansteelandt (2010); Imai et al. (2010); Lange and Hansen (2011); VanderWeele (2011)). The mediation formula suggests a generic approach for mediation analysis, but has a number of important limitations. First, the way to compute natural direct and indirect effects, and their standard errors, can differ substantially depending on the model for the mediator and outcome. Second, even simple models for the mediator and outcome (e.g. a linear model for the mediator and logistic regression model for the outcome) tend to produce complex expressions of natural direct and indirect effects. This can make results difficult to report. In addition, it can make testing for the presence of a direct or indirect effect essentially impossible, as it generally turns out difficult to choose models for the outcome and mediator that result in a natural direct effect of zero at all covariate levels.

van der Laan and Petersen (2008) accommodated this by directly modeling the natural direct effect of interest. More precisely, because application of the mediation formula to even simple models for the mediator and outcome typically results in complicated models for the natural

direct and indirect effect, they instead choose to a priori postulate a parsimonious model structure for the natural direct effect. Results thereby become simpler for reporting and interesting hypotheses concerning these effects become easier to test. van der Laan and Petersen (2008) proposed doubly robust estimators for the direct effect parameters indexing their model, which require correct specification of a model for the distribution of the mediator (given exposure and confounders), and for either the distribution of the exposure (given confounders) or the expected outcome (given mediator, exposure and confounders). Tchetgen Tchetgen and Shpitser (2011) proposed estimators with greater robustness against model misspecification, which require an arbitrary 2 out of the 3 models for the exposure, mediator and outcome to be correctly specified. These approaches are very elegant and appealing; arguably, their relative complexity may be a barrier to routine application.

The goal of this article is to enhance the accessibility of methods for causal mediation analysis based on natural direct and indirect effects. We will attempt this by drastically widening the scope of models (and hence outcome types) which can be dealt with, and by proposing flexible estimation strategies for natural direct and indirect effects that are easy to perform with standard statistical software. In particular, we propose a simple regression mean imputation approach which relies on correct specification of a model for the outcome. In addition, we propose a doubly robust imputation approach, which relaxes the assumptions of the simple imputation estimator by continuing to be valid when the outcome model is misspecified, provided that models for both mediator and exposure are correctly specified. We further show that the doubly robust imputation approach can be viewed as a compromise between the simple imputation approach and an inverse probability weighting approach previously considered by Hong (2010) and Lange et al. (2011). We give a theoretical discussion of the properties of these estimators and assess their finite-sample performance through a simulation study and through the analysis of the WHO LARES study on the association between residence in a damp and moldy dwelling and the risk of depression.

5.2 Natural direct and indirect effects

Let A be the observed exposure of interest, M the mediator, C a set of baseline covariates and Y the outcome. Variables are allowed to be of any type e.g. continuous, binary, categorical, or survival. Then, as in Robins and Greenland (1992) and Pearl (2001) we will describe direct and indirect effects in terms of so-called nested counterfactuals, $Y(a, M(a^*))$, denoting the outcome that would have been observed if A were set to a and M to the value it would have taken if A were set to a^*. In particular, we will compare $Y(a, M(a^*))$ with $Y(a^*, M(a^*))$ to obtain a measure of the direct effect of changing the exposure from a to a^*. Such comparison can for instance be made in terms of an average difference within levels of covariates, $E[Y(a, M(a^*)) - Y(a^*, M(a^*))|C]$, as a risk ratio, $P[Y(a, M(a^*)) = 1|C]/P[Y(a^*, M(a^*)) = 1|C]$, etc. We will refer to such contrasts as natural direct effects. Likewise we will compare $Y(a^*, M(a))$ with $Y(a^*, M(a^*))$ to obtain a measure of the indirect effect, referred to as a natural indirect effect.

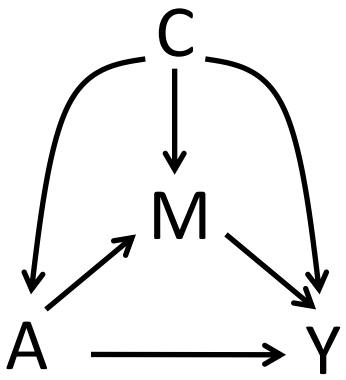

Figure 5.1: Causal diagram.

Throughout the article, we will assume that the causal assumptions underlying the causal diagram of Figure 5.1 hold. In particular, we will assume that the same set of covariates C is sufficient to control for confounding of the associations between exposure and outcome, exposure and mediator, and mediator and outcome. Specifically, we thus assume that there are no variables that are effects of exposure and that confound the mediator-outcome relationship. A

formal description of these assumptions in terms of counterfactuals is for instance given in VanderWeele and Vansteelandt (2009). Under these assumptions, it follows from Pearl (2001, 2011) that $E[Y(a, M(a^*))|C = c]$ can be calculated using the mediation formula:

$$\sum_m E[Y|A = a, M = m, C = c]P(M = m|A = a^*, C = c). \tag{5.1}$$

This corresponds to estimating the mean value of the outcome in each stratum defined by mediator and confounders among the individuals with exposure a, but weighting these by the likelihood of each mediator value among individuals with exposure a^*. On the basis of this, natural direct and indirect effects can be obtained. For instance, if the outcome Y and mediator M obey linear models, i.e. $E[Y|A = a, M = m, C = c] = \alpha_0 + \alpha_1 a + \alpha_2 m + \alpha_3 c$ and $E[M|A = a, C = c] = \beta_0 + \beta_1 a + \beta_2 c$, then equation (5.1) greatly simplifies and the natural direct and indirect effects are captured by $E[Y(a, M(a^*)) - Y(a^*, M(a^*))|C] = \alpha_1(a - a^*)$ and $E[Y(a^*, M(a)) - Y(a^*, M(a^*))|C] = \alpha_2\beta_1(a - a^*)$ (VanderWeele and Vansteelandt, 2009). However, equation (5.1) lends itself less ideally to non-linear models, where it tends to generate complicated expressions of natural direct and indirect effects (see e.g. VanderWeele and Vansteelandt (2010) who find that the combination of a logistic model for the outcome and a linear model for the mediator tends to generate complicated natural direct effect expressions depending on covariates, even when the outcome is rare).

5.3 Simultaneous estimation of natural direct and indirect effects

5.3.1 Natural effect models

In this paper, as in Lange et al. (2011), we focus on conditional mean models for nested counterfactuals $Y(a, M(a^*))$:

$$g\left[E\left\{Y(a, M(a^*))|C\right\}\right] = \beta' W(a, a^*, C),$$

where $g(.)$ is a canonical link function (e.g. the identity or logit link) and $W(a, a^*, C)$ is a known vector with components that may depend on a, a^* and C. We term such models 'natural effect models' because they encode the natural direct and indirect effects of interest. For instance, in model

$$E[Y(a, M(a^*))|C] = \beta_0 + \beta_1 a + \beta_2 a^* + \beta_3 C, \tag{5.2}$$

$\beta_1(a - a^*)$ captures a natural direct effect and $\beta_2(a - a^*)$ captures a natural indirect effect. Under model

$$\text{logit}\left[E\left\{Y(a, M(a^*))|C\right\}\right] = \beta_0 + \beta_1 a + \beta_2 a^* + \beta_3 C + \beta_4 a \cdot a^* + \beta_5 a \cdot C + \beta_6 a^* \cdot C, \tag{5.3}$$

we find that the natural direct effect odds ratio

$$\frac{\text{odds}\left\{Y(a, M(a^*)) = 1|C\right\}}{\text{odds}\left\{Y(a^*, M(a^*)) = 1|C\right\}} = \exp\left\{(\beta_1 + \beta_4 a^* + \beta_5 C)(a - a^*)\right\}$$

and that the natural indirect effect odds ratio

$$\frac{\text{odds}\left\{Y(a, M(a)) = 1|C\right\}}{\text{odds}\left\{Y(a, M(a^*)) = 1|C\right\}} = \exp\left\{(\beta_2 + \beta_4 a + \beta_6 C)(a - a^*)\right\}.$$

Their product measures the total effect: $\text{odds}\left\{Y(a) = 1|C\right\} / \text{odds}\left\{Y(a^*) = 1|C\right\}$ (VanderWeele and Vansteelandt, 2010).

We note that the considered class of natural effect models is richer than the classes of models considered by van der Laan and Petersen (2008) and Tchetgen Tchetgen and Shpitser (2011). First, it is richer in the sense that it simultaneously describes natural direct and indirect effects, rather than separately. Second, it is richer by not being limited to the identity and log links. It thus in particular enables the analysis of dichotomous outcomes, as in the motivating application of Section 5.5.

In the following sections, we will discuss three estimation strategies for the parameters index-

ing the natural effect model. For simplicity, we will focus on dichotomous exposures A taking values 0 and 1. We will expand on the analysis of general exposures in the discussion section.

5.3.2 Simple IPW-estimator

Generalizing work by Hong (2010), Lange et al. (2011) showed that consistent estimators of the parameters indexing natural effect models (i.e., estimators which approach the targeted population value as the sample size goes to infinity) can be obtained by the following approach, which is easy to perform in routine software packages:

1. Fit a suitable model for the mediator conditional on exposure and covariates using the original data set.

2. Create a new data set by repeating each observation in the original data set twice and including an additional variable A^*, which is equal to the original exposure for the first replication and equal to the opposite of the actual exposure for the second replication. In addition, add an id-variable to memorize which pseudo observations originate from the same subject.

3. Compute a weight
$$\frac{f(M = m|A = A^*, C)}{f(M = m|A = A, C)}$$
in each line of the new data set using predict-functionality on the fitted model.

4. Fit the natural effect model of interest by regressing the outcome on A, A^* and C on the basis of the expanded data set, weighting the data by the weights from the previous step. Lange et al. (2011) show that, provided the model for the mediator is fitted using a standard maximum likelihood procedure, conservative confidence intervals can be obtained as the estimate of the natural direct or indirect effect plus/minus 1.96 times a robust standard error, which can be obtained using software for generalized estimating equations.

The Inverse Probability Weighted (IPW) estimator is appealing for routine use, but its simplicity may come at a cost. First, the IPW-estimator does not make efficient use of the information in the data in the sense that more efficient estimators can be constructed under the joint model defined by the natural effect model together with the model for the mediator. Second, the requirement of full specification of the conditional mediator distribution can be prohibitive for quantitative mediators. Third, when the exposure and/or covariates are strongly predictive of the mediator, or when the mediator is quantitative, then the inverse weights may easily become large for some subjects, making them very influential in the analysis. This can lead to an inflation of the standard errors and, more importantly, to unstable estimators with an important degree of finite sample bias. In particular, it may make the IPW approach of more limited use for the analysis of continuous mediators. In the following sections, we propose alternative estimators designed to overcome these limitations.

5.3.3 Simple (regression mean) imputation estimator

The nested counterfactual $Y(a^*, M(a))$ is only observable when a^* equals a and, in addition, a corresponds to the observed exposure level A. When a^* differs from a, then it can still be predicted as $E(Y|A = a^*, M, C)$ under a model for the outcome. This can be seen upon noting that $M(a)$ equals M among subjects with a equalling the observed value A. This thus gives rise to the following regression mean imputation approach:

1. Fit a suitable model for the outcome conditional on exposure, mediator and baseline variables using the original data set.

2. Create a new data set by repeating each observation in the original data set twice and including an additional variable A^*, which is equal to the original exposure for the first replication and equal to the opposite of the actual exposure for the second replication.

3. Predict the nested counterfactual $Y(a^*, M(a))$ as Y in those lines of the database where A equals A^* and as the expected value $E(Y|A = A^*, M, C)$ on the remaining lines, using

predict-functionality on the fitted outcome model.

4. Fit the natural effect model of interest by regressing the outcome on A, A^* and C on the basis of the expanded data set. Standard errors and confidence intervals can be obtained using the bootstrap (including steps 1-4).

The (regression mean) imputation estimator is appealing for routine use because of its simplicity and avoidance of inverse probability weighting. However, its simplicity may be somewhat deceptive in that the difficulty in working with the imputation estimator - as with multiple imputation estimators for incomplete data analyses - is finding models for the outcome that are congenial with the natural effects model of interest. Suppose for instance that the model for the outcome were of the form

$$E(Y|M, A, C) = \alpha_0 + \alpha_1 M + \alpha_2 A + \alpha_3 C$$

and the natural effects model of interest were

$$E[Y(a, M(a^*))|C] = \beta_0 + \beta_1 a + \beta_2 a^* + \beta_3 a a^* + \beta_4 C.$$

Then, under data generating models where M is linear in A, the outcome model would preclude the existence of interactions between a and a^* and thus bias β_3 towards zero. With concern for such interactions, the user should therefore work with outcome models which are sufficiently flexible not to preclude the existence of such interactions. As a guideline, we therefore recommend that the outcome model should at least reflect the structure of the natural effect model, but with a^* substituted by M. Thus, given the above natural effects model, the outcome model should at least contain main effects in A, M and C, as well as the product term $A \times M$.

This concern for the imputation model (i.e., the outcome model) failing to reflect the structure of the natural effects model gets exacerbated in non-linear models, where it becomes difficult

or impossible to guarantee coherent model specifications. Noteworthy is that the imputation estimator - like other imputation estimators (Tan, 2007; Vansteelandt et al., 2010) - may also be more prone to model extrapolation, primarily in settings where A shows strong associations with either M or C.

5.3.4 Doubly robust estimator

The foregoing concerns have lead us to consider a third estimator, which could be viewed as a compromise between both previous approaches:

1. Fit a suitable model for the mediator conditional on exposure and baseline variables using the original data set.

2. Fit a suitable model for the exposure conditional on mediator and baseline variables using the original data set.

3. Fit a suitable generalized linear model for the outcome conditional on exposure, mediator and baseline variables using the standard maximum likelihood procedure, but using the weights
$$\frac{f(A=1|C)}{f(A=0|C)}\frac{f(M|A=1,C)}{f(M|A=0,C)}$$
for the unexposed ($A=0$) and
$$\frac{f(A=0|C)}{f(A=1|C)}\frac{f(M|A=0,C)}{f(M|A=1,C)}$$
for the exposed ($A=1$).

4. Create a new data set by repeating each observation in the original data set twice and including an additional variable A^*, which is equal to the original exposure for the first replication and equal to the opposite of the actual exposure for the second replication.

5. Predict the nested counterfactual $Y(a^*, M(a))$ as Y in those lines of the database where

A equals A^* and as the expected value $E(Y|A = A^*, M, C)$ on the remaining lines, using predict-functionality on the fitted models obtained in step 3.

6. Fit the natural effect model of interest by regressing the outcome on A, A^* and C on the basis of the expanded data set. Standard errors and confidence intervals can be obtained using the bootstrap (including steps 1-6).

In the Appendix, we show that the resulting estimator is a doubly robust estimator of the parameters indexing the natural effect model. More precisely, it is a consistent estimator of these parameters when the natural effect model holds and, in addition, either (1) the conditional expectation of the outcome, given the exposure and confounders, is correctly specified; or (2) the conditional distribution of the mediator, given the exposure and confounders, and of the exposure, given confounders, are correctly specified. This double robustness is theoretically appealing because it implies that incongeniality of the outcome model and the natural effect model is not of concern when the exposure and mediator models are reasonably well specified. In addition, it implies that truncation of extreme inverse probability weights - which could be viewed as a deliberate form of model misspecification - is justified when the outcome model is reasonably well specified. Note in particular that the simple imputation estimator is attained upon truncating all weights in the doubly robust estimator to 1.

Table 5.1 summarizes the properties of the considered estimators by indicating under what models they produce consistent estimators.

Estimator	**Weighting**	**Imputation**	**Model for**		
			A	M	Y
IPW	Yes	No		X	
IMP	No	Yes			X
DR	Yes	Yes	X	X	X

Table 5.1: Properties of the simple IPW-estimator (IPW), the simple imputation estimator (IMP) and the doubly robust (DR) estimator. Crosses (X) indicate what models need to be correctly specified for the estimator to be consistent.

5.4 Simulation study

To assess the performance of the three proposed natural direct and indirect effects estimators, a simulation study with 1000 runs for data sets of 500 observations was performed. We evaluated all proposed estimators as well as the more traditional estimators based on the product of coefficient method and difference in coefficient methods for the indirect effect. Standard errors were obtained using the nonparametric bootstrap (1000 resamples) and the coverage of 95% standard normal approximation bootstrap confidence intervals is reported. In all simulation experiments, the natural effects model of interest was correctly specified.

5.4.1 Continuous outcome and binary mediator

The first setting focuses on a continuous outcome and binary mediator. In particular, C was generated from a standard normal distribution. A binary exposure A and mediator M were drawn from a binomial distribution with $P(A = 1|C) = \text{expit}(0.25 + 0.5C)$ and $P(M = 1|A, C) \equiv \pi(A, C) = \text{expit}(-0.5 + A + C)$. The outcome Y was generated from a normal distribution with mean

$$E(Y|A = a, M = 0, C = c) = \beta_0 + \beta_1 a + \beta_3 c - \frac{\beta_2 \pi(0, c)}{\pi(1, c) - \pi(0, c)}$$
$$E(Y|A = a, M = 1, C = c) = \beta_0 + \beta_1 a + \beta_3 c + \frac{\beta_2 \{1 - \pi(0, c)\}}{\pi(1, c) - \pi(0, c)}$$

and variance equal to 2. In the appendix we show that these models were chosen to result in a simple natural effects model of the form $E[Y(a, M(a^*))|C] = \beta_0 + \beta_1 a + \beta_2 a^* + \beta_3 c$, which can be verified upon applying the mediation formula.

Table 5.2 and 5.3 show simulation results for a settings with natural direct effect $\beta_1 = 0.5$ and natural indirect effect $\beta_2 = 0.1$. In the first simulation experiment (Table 5.2), we fitted correctly specified models for the exposure and mediator, but misspecified the model for the outcome which we chose to be linear with main effects in A, M and C. The weights corre-

	Estimator	Estimate	SE	Variance	bias^2	MSE	Coverage
DE	Regression	0.49	0.10	0.011	1 10^{-4}	0.011	0.94
	IPW	0.50	0.10	0.012	0	0.012	0.94
	Imputation	0.49	0.10	0.011	1 10^{-4}	0.011	0.93
	DR	0.51	0.11	0.015	1 10^{-4}	0.015	0.94
IE	Product	0.48	0.15	0.022	0.002	0.024	0.18
	Difference	0.11	0.040	0.002	1 10^{-4}	0.002	0.96
	IPW	0.10	0.042	0.002	0	0.002	0.94
	Imputation	0.11	0.040	0.002	1 10^{-4}	0.002	0.96
	DR	0.10	0.075	0.007	0	0.007	0.95

Table 5.2: Results from setting 1, simulation experiment 1: direct effect (DE) = 0.5, indirect effect (IE) = 0.1, correctly specified model for A and M.

sponding to the IPW (min. 0.4, median 1, max. 2.5) and the doubly robust estimator (min. 0.2, median 0.6, max. 5) were very stable. In the second experiment (Table 3) we additionally misspecified the models for the mediator by using the probit link rather than the logit link.

In the first experiment, all estimators perform well, with the exception of the doubly robust estimator of the indirect effect, which has large variance due to the relatively less stable weights. The regression mean imputation estimator is identical to the standard (difference of coefficient) estimator based on regression adjustment for the mediator. However, a large bias is found in the product of coefficient method for the indirect effect as a result of this strategy being essentially limited to linear models for both outcome and mediator. Interestingly, neither the regression mean imputation estimator nor the doubly robust estimator of the indirect effect are biased, in spite of misspecification of the outcome model. Similar conclusions are obtained in the second experiment in spite of misspecification of the mediator model, which harmfully affected only the variance of the doubly robust estimator.

5.4.2 Binary outcome and continuous mediator

The second setting focuses on a binary outcome and continuous mediator. In particular, a binary exposure A was drawn from a binomial distribution with $P(A = 1|C) = \text{expit}(\gamma_0 + \gamma_1 C)$

	Estimator	Estimate	SE	Variance	bias^2	MSE	Coverage
DE	Regression	0.49	0.10	0.011	1 10^{-4}	0.011	0.94
	IPW	0.49	0.11	0.015	1 10^{-4}	0.012	0.94
	Imputation	0.49	0.10	0.011	1 10^{-4}	0.011	0.93
	DR	0.51	0.12	0.017	1 10^{-4}	0.017	0.93
IE	Product	0.29	0.089	0.008	0.0008	0.008	0.44
	Difference	0.11	0.040	0.002	1 10^{-4}	0.002	0.96
	IPW	0.10	0.042	0.002	0	0.002	0.95
	Imputation	0.11	0.040	0.002	1 10^{-4}	0.002	0.96
	DR	0.11	0.093	0.010	1 10^{-4}	0.01	0.95

Table 5.3: Results from setting 1, simulation experiment 2: direct effect (DE) = 0.5, indirect effect (IE) = 0.1, correctly specified model for A but misspecified model for M.

with $\gamma = (0.25, -0.5)$. The continuous mediator M was drawn from a normal distribution with $E(M|A, C) = \alpha_0 + \alpha_1 A + \alpha_2 C$ with $\alpha = (1, 3, -5)$ and variance $\sigma^2 = 1$. The binary outcome Y was drawn from a binomial distribution with $P(Y = 1|A, M, C) = \Phi(\beta_0 + \beta_1 A + \beta_2 M + \beta_3 C)$, with $\beta = (0.5, 0.1, -0.05, 0.5)$ where $\Phi(.)$ refers to the cumulative standard normal distribution function. It is can be verified upon applying the mediation formula that this choice of models entails a natural effects model of the form

$$P[Y(a, M(a^*)) = 1|C] = \Phi\left(\theta'_0 + \theta'_1 a + \theta'_2 a^* + \theta'_2 C\right)$$

where

$$\theta'_0 = \frac{\theta_0 + \theta_2 \alpha_0}{\sqrt{1 + \beta_2^2 \sigma^2}}$$
$$\theta_k = \frac{\theta_k}{\sqrt{1 + \beta_2^2 \sigma^2}},$$

for $k = 1, 2, 3$.

The weights corresponding to the IPW estimator (min. 1.2 10^{-6}, median 1, max. 391) and the doubly robust estimator (min. 9.1 10^{-7}, median 0.01, max. 455) were very unstable. Because of unstable weights, estimators based on inverse probability weights are presented both

	Estimator	Estimate	SE	Variance	bias^2	MSE	Coverage
DE	Regression	0.099	0.23	0.055	1 10^{-6}	0.23	0.95
	IPW	0.093	6.4 10^{14}	0.33	4.9 10^{-5}	0.33	1
	IPW$_{trunc}$	0.028	0.17	0.061	5.2 10^{-3}	0.17	0.79
	Imputation	0.098	0.23	0.055	4 10^{-6}	0.055	0.95
	DR	0.11	2.23	0.61	1 10^{-4}	0.61	1
	DR$_{trunc}$	0.10	0.35	0.20	0	0.20	0.88
IE	Product	-0.15	0.19	0.037	1 10^{-4}	0.037	0.95
	Difference	-0.15	0.19	0.037	1 10^{-4}	0.037	0.95
	IPW	-0.13	6.5 10^{14}	0.31	4 10^{-4}	0.31	1
	IPW$_{trunc}$	-0.074	0.087	0.041	5.8 10^{-3}	0.041	0.58
	Imputation	-0.14	0.19	0.037	1 10^{-4}	0.037	0.95
	DR	-0.16	2.22	0.56	1 10^{-4}	0.56	1
	DR$_{trunc}$	-0.15	0.32	0.19	0	0.19	0.89

Table 5.4: Results from setting 2: all models correctly specified with and without truncated weights. DE and IE are simulated to be 0.1 and -0.15 respectively

with and without truncation of the weights at a maximum of 10 (max) and a minimum of 0.1. Results in table 5.4 show that the IPW-estimator has little bias which may be due to the large sample size, but that it is very inefficient as a result of extreme weights. Truncating the weights pays off in terms of efficiency, but at the expense of a large bias because weight truncation can be viewed as a form of misspecification of the model for the mediator. This is not the case for the doubly robust estimator as a result of the additional robustness of this estimator against model misspecification. However, this estimator does not compete well with the regression mean imputation estimator, which is much more efficient here. The standard regression approach performs very well (not surprisingly, because it uses the data-generating models and because $\sqrt{1+\beta_2^2\sigma^2} \approx 1$).

With concern for model incongeniality, we additionally evaluated a setting where we chose the model for the outcome to be a logistic regression model. The results in Table 5.5 show that, interestingly, this does not affect the adequate performance of the regression mean imputation estimator.

	Estimator	Estimate	SE	Variance	bias^2	MSE	Coverage
DE	Imputation	0.095	0.23	0.055	4 10^{-6}	0.055	0.94
	DR	0.11	2.3	0.59	1 10^{-4}	0.59	1
	DR$_{trunc}$	0.10	0.35	0.19	0	0.19	0.88
IE	Imputation	-0.14	0.19	0.037	1 10^{-4}	0.037	0.95
	DR	-0.15	2.3	0.55	0	0.56	1
	DR$_{trunc}$	-0.14	0.32	0.19	0	0.19	0.89

Table 5.5: Results from an experiment with a misspecified model for the outcome. DE and IE are simulated to be 0.1 and -0.15 respectively

Table 5.6 shows additional results from a setting with very strong mediator-outcome relationship ($\beta_2 = 1$) resulting in $\sqrt{1 + \beta_2^2\sigma^2} = 1.41$. The corresponding direct and indirect effect were equal 0.35 en 2.12 respectively. The weights corresponding to the IPW estimator (min. 2.5 10^{-6}, median 1, max. 215) and the doubly robust estimator (min. 1.8 10^{-8}, median 0.01, max. 208) were again very unstable.

	Estimator	Estimate	Variance	bias^2	MSE
DE	Regression	0.52	0.29	0.03	0.32
	IPW	1.15	0.64	0.64	1.28
	IPW$_{trunc}$	1.53	0.30	1.38	1.68
	DR	0.51	0.88	0.02	0.90
	DR$_{trunc}$	0.42	0.41	0.004	0.42
	Imputation	0.38	0.15	0.001	0.15
IE	Product	3.19	0.41	1.14	1.55
	Difference	2.03	0.12	0.01	0.13
	IPW	1.73	0.45	0.15	0.60
	IPW$_{trunc}$	1.06	0.08	1.14	1.21
	DR	2.15	0.41	0.001	0.41
	DR$_{trunc}$	2.22	0.20	0.01	0.21
	Imputation	2.18	0.07	0.003	0.07

Table 5.6: Results from the setting with a strong mediator-outcome relationship. DE and IE are simulated to be 0.35 and 2.12 respectively

Due to noncollapsibility, the regression estimator based on the product of coefficients is severely biased. The simple IPW estimator for the direct and indirect effect has now a large bias

(even without truncation of the weights) and is very inefficient. The doubly robust estimator performs pretty well after trunctation of the extreme weights. But again, in terms of the MSE, the imputation based estimator performs the best and actually extremely well compared to the other estimators.

5.5 Data analysis

We re-analyze data of 5882 adult respondents from the WHO-LARES study (Shenassa et al., 2007; VanderWeele and Vansteelandt, 2010), which investigated the association between residence in a damp and moldy dwelling and the risk of depression. Interest lies in whether the effect of residence in a damp and moldy dwelling (A) on depression (Y) is mediated by perception of control over one's home or mold-related physical illness, measured on a 5-point Likert scale (reverse coded) (M). A large set of potential confounders C is available on city of residence, survey respondent characteristics (age, gender, marital status, education, employment, smoking, and environmental tobacco smoke at home) and housing characteristics (ownership, size, tenure, crowding, ventilation, natural light, and heating).

Tables 5.10, 5.11 and 5.12 in the Appendix show the results of a logistic regression model for the outcome, mediator and exposure. They show differences in depression rates and in perception of control over one's home or mold-related physical illness between different cities of residence, by respondent characteristics age, gender marital status, education and environmental tobacco smoke, and by housing characteristics crowding and natural light.

For simplicity, we first focus on the natural effects model given by

$$\text{logit}P\left\{Y(a, M(a^*)) = 1|C\right\} = \beta_0 + \beta_1 a + \beta_2 a^* + \beta_3 C.$$

The estimates of the natural direct and indirect odds ratios corresponding to the different estimators based on models as in Table 5.10, 5.11 and 5.12 are given in Table 5.7. Both the IPW-estimate and the imputation-based estimate suggest that the natural direct and indirect effect of

	Estimator	OR	95% CI	P-value
Direct effect	Regression	1.36	1.10 to 1.68	0.0048
	IPW	1.35	1.10 to 1.67	0.0049
	Imputation	1.35	1.10 to 1.67	0.0048
	DR	1.32	1.05 to 1.65	0.018
Indirect effect	Product	1.05	1.03 to 1.08	< 0.0001
	Difference	1.05	1.03 to 1.08	0.0001
	IPW	1.05	1.03 to 1.08	< 0.0001
	Imputation	1.06	1.03 to 1.08	0.0001
	DR	1.07	1.01 to 1.13	0.021

Table 5.7: Estimates of the natural direct and indirect effect corresponding to different estimation procedures.

the presence of dampness or mold exposure on the odds of depression amounts to odds ratios $e^{0.30} = 1.35$ (95% CI 1.10 to 1.67, P 0.005) and $e^{0.053} = 1.06$ (95% CI 1.03 to 1.08, P < 0.001), respectively. That is, if none of the residents were to reside in a damp and moldy dwelling, then the effect of residence in damp and moldy dwelling would be to increase the odds of depression with 35% if they were to maintain their perception of control over one's home. Furthermore, if all of the residents were to reside in a damp and moldy dwelling, then the effect of changing their perception of control over one's home to what it would be if they were not to reside in such a residence, would be to reduce the odds of depression with 6%. Note that conservative standard errors and confidence intervals based on the IPW-estimate can be directly obtained from software for generalized estimating equations and were found to be exactly the same here as the bootstrap confidence intervals. The DR-estimates of the natural direct and indirect odds ratio differ slightly: 1.32 (95% CI 1.05 to 1.65, P 0.018) and 1.07 (95% CI 1.01 to 1.13, P 0.021), respectively. The weights corresponding to the IPW estimator (min. 0.46, median 1, max. 2.15) were more stable that those corresponding to the DR estimator (min. 0.026, median 0.60, max. 19.1).

We finally fitted the natural effects model given by

$$\text{logit}P\left\{Y(a, M(a^*)) = 1|C\right\} = \beta_0 + \beta_1 a + \beta_2 a^* + \beta_3 C + \beta_4 aC + \beta_5 a^* C,$$

and found, using the IPW estimator, the magnitude of the natural direct effect to depend on the number of residents per room and the strength of the natural indirect effect to depend on the amount of natural light (see Table 5.8). The interaction of the natural direct effect with the number of residents per room was also found by the imputation and the doubly robust estimator (see Table 5.9). The evidence for the natural indirect effects interaction disappeared when using the regression imputation estimator and the doubly robust estimator based on previous model for the outcome (see Table 9 for corresponding p-values). This is because this outcome model is not congenial with the natural effect model in the sense that it excludes such interactions. Upon extending the outcome model by including interactions of A with the number of residents per room and of M with the amount of natural light , the p-values of corresponding interactions decreased a bit but were not significant due to inefficient parameter estimates. This underscores the importance of choosing the model for the outcome in a way that is congenial with the final analysis model.

Interaction	**Level**	**OR**	**95% CI**	**P-value**
Direct effect				
Crowding	<0.5	2.04	1.40 to 2.98	0.0002
	0.51-1	1.06	0.80 to 1.41	0.68
	>1 per room	1.35	0.91 to 2.02	0.13
Indirect effect				
Light	enough	1.10	1.06 to 1.15	<0.0001
	not enough	1.04	1.01 to 1.06	0.005

Table 5.8: IPW estimates of the natural direct and indirect effect with confounder-interactions.

	C	Level	IPW	Imputation	DR
Direct effect	Crowding	< 0.5	0.006	0.006	0.003
	Crowding	0.51 to 1	0.14	0.13	0.12
Indirect effect	Light	*Enough*	0.019	0.17	0.28

Table 5.9: P-values for the interactions of direct and indirect effect with baseline covariates C

5.6 Discussion

In this paper, we have focused on so-called natural effect models, whose parameters encode the natural indirect effect of an exposure on an outcome, through a given mediator, and the remaining natural direct effect. We have reviewed and proposed estimators for the parameters indexing these models based on inverse probability weighting, or imputation, or a combination of both. For pedagogic purposes, we have explained the approach for generalized linear natural effect models, including linear, loglinear and logistic models, although the IPW-estimator and the simple imputation estimator extend to other types of models, such as parametric survival models. The proposed doubly robust estimator is essentially confined to generalized linear natural effect models, but the results in the Appendix show how such estimators may be obtained for a wider class of models. Noteworthy is that confounding adjustment in all considered approaches is based on regression adjustment, and thus that the natural effect models must include all relevant confounders, even when their association with the outcome is not of immediate interest.

All proposed estimators are easy to obtain using standard statistical software, but differ in terms of their properties. The simple IPW-estimator does not make efficient use of the information in the data and can suffer instability when some subjects in the analysis get assigned large weights. While weight truncation may improve stability, this may come at the expense of bias. The simple imputation estimator avoids inverse weighting and thereby tends to give more precise estimators. However, it requires more care in choosing an adequate imputation model and may overstate the precision of the resulting estimators by ignoring extrapolation uncertainty. From a theoretical point of view, the doubly robust imputation estimator seems a useful compromise: it is valid

when either the imputation model, or both the exposure and mediator models are correctly specified. In large samples, it may thus turn out to be less sensitive to bias resulting from truncation of the inverse probability weights, as well as to bias resulting from the lack of a coherent model specification. However, simulations studies and the data analysis have demonstrated that its practical utility may be more limited because the inverse probability weights that it relies upon tend to be much less stable than those employed by the simple IPW-estimator. Furthermore, the asserted property of double robustness may be somewhat illusory (Robins and Rotnitzky, 2001) because correct specification of the exposure and mediator model may anyway be essential in the face of incoherent model specification. In this respect, the estimators of van der Laan and Petersen (2008) and Tchetgen Tchetgen and Shpitser (2011) promise a more guaranteed robustness against model misspecification, but rely on similar inverse probability weights and will thus likely suffer the same instability.

Our recommendation is to use the simple imputation estimator for routine analysis. We make this recommendation in view of the foregoing discussion, the empirical simulation findings, and the fact that the concerns about model incongeniality may be relatively lenient if one is careful to make the imputation model sufficiently flexible - as also confirmed in the simulation studies. Note for instance that concerns about model incongeniality are much more conspicuous in the widely adopted multiple imputation strategies for the analysis of incomplete data, where the imputation model typically not just refers to the outcome distribution, but to the high-dimensional distribution of outcomes and covariates. We do believe that the simple IPW-estimator and the doubly robust estimator are valuable, but primarily to confirm that similar results are obtained upon relying on different models.

Our development has been confined to dichotomous exposures. For continuous exposures, we recommend changing the second step of the imputation algorithm as follows: create a new data set by repeating each observation in the original data set K times and including an additional

variable A^*, which is equal to the original exposure for the first replication and equal to a (different) random draw from the conditional exposure distribution, given C, for all $K-1$ remaining replications. While K is ideally as large as possible, computational resources will restrict the maximum number that can be chosen in any given application. In future work, we will investigate whether this recommendation can be improved with a view towards both statistical efficiency and computation time.

In summary, we have proposed and evaluated procedures for mediation analysis that are easy to perform with standard software. We hope that the flexibility of the procedures for handling arbitrary mediators and outcomes, as well as their simplicity, will aid to make inference for natural direct and indirect effects accessible to a wider audience.

Appendix

Derivation of the doubly robust estimator

In this section, we will explain how to obtain doubly robust estimators of the parameters indexing the more general natural effects model

$$E\left\{Y(a, M(a^*))|C\right\} = m(a, a^*, C; \beta),$$

where $m(a, a^*, C; \beta)$ is a known function, smooth in an unknown finite-dimensional parameter β.

We start with a sketch of the derivation of the set of influence functions (Newey, 1990) for β in the above model, which is closely related to the derivation given in Tchetgen and Schpitser (2011). First note that

$$m(a, a^*, C; \beta) = \int E(Y|A = a, M = m, C) f(M = m|A = a^*, C) dm,$$

and let a and a^* be fixed values for now. Taking derivatives along one-dimensional parametric submodels, indexed by t, we obtain

$$\begin{aligned} 0 = \int \int & y f(Y = y|A = a, M = m, C) f(M = m|A = a^*, C) \\ & \times \left\{S_y(y|A = a, M = m, C) + S_m(M = m|A = a^*, C)\right\} dydm, \end{aligned}$$

where $S_y(y|A = a, M = m, C)$ and $S_m(M = m|A = a^*, C)$ are the scores corresponding to the components $f(Y = y|A = a, M = m, C)$ and $f(M = m|A = a^*, C)$, respectively. Note

that

$$
\begin{aligned}
&\int\int yf(Y=y|A=a,M=m,C)f(M=m|A=a^*,C)S_y(y|A=a,M=m,C)dydm \\
&=\int\int\int y\frac{I(A=a)}{f(A=a|C)}\frac{f(M=m|A=a^*,C)}{f(M=m|A=a^*,C)}S_y(y|A=a,M=m,C) \\
&\qquad\times f(Y=y,M=m,A|C)dydmdA \\
&=\int\int\int\{y-E(Y|A=a,M=m,C)\}\frac{I(A=a)}{f(A=a|C)}\frac{f(M=m|A=a^*,C)}{f(M=m|A=a,C)} \\
&\qquad\times S_y(y|A=a,M=m,C)f(Y=y,M=m,A|C)dydmdA \\
&=\int\int\int\{y-E(Y|A=a,M=m,C)\}\frac{I(A=a)}{f(A=a|C)}\frac{f(M=m|A=a^*,C)}{f(M=m|A=a,C)} \\
&\qquad\times\{S_y(y|A=a,M=m,C)+S_m(M=m|A=a,C)+S_a(A=a|C)\} \\
&\qquad\times f(Y=y,M=m,A|C)dydmdA,
\end{aligned}
$$

and

$$
\begin{aligned}
&\int E(Y|A=a,M=m,C)f(M=m|A=a^*,C)S_m(M=m|A=a^*,C)dm \\
&=\int\int E(Y|A=a,M=m,C)\frac{I(A=a^*)}{f(A=a^*|C)}S_m(M=m|A=a^*,C) \\
&\qquad\times f(M=m,A|C)dmdA \\
&=\int\int\{E(Y|A=a,M=m,C)-m(a,a^*,C;\beta)\}\frac{I(A=a^*)}{f(A=a^*|C)}S_m(M=m|A=a^*,C) \\
&\qquad\times f(M=m,A|C)dmdA \\
&=\int\int\int\{E(Y|A=a,M=m,C)-m(a,a^*,C;\beta)\}\frac{I(A=a^*)}{f(A=a^*|C)} \\
&\qquad\times\{S_y(y|A=a,M=m,C)+S_m(M=m|A=a,C)+S_a(A=a|C)\} \\
&\qquad\times f(Y=y,M=m,A|C)dydmdA.
\end{aligned}
$$

We conclude from both these expressions that for each one-dimensional parametric submodel,

the score $U = S_y(y|A=a,M,C) + S_m(M|A=a,C) + S_a(A=a|C)$ satisfies

$$0 = E\left[U\left\{\{Y - E(Y|A=a,M,C)\}\frac{I(A=a)}{f(A=a|C)}\frac{f(M|A=a^*,C)}{f(M|A=a,C)} + \{E(Y|A=a,M,C) - m(a,a^*,C;\beta)\}\frac{I(A=a^*)}{f(A=a^*|C)}\right\}|C\right],$$

for each a, a^*. It is now immediate from Theorem 2.4 in Newey (1990) that all influence functions under the mediation model for fixed a and a^* are given by

$$d_{a,a^*}(C)\left[\{Y - E(Y|A=a,M,C)\}\frac{I(A=a)}{f(A=a|C)}\frac{f(M|A=a^*,C)}{f(M|A=a,C)} + \{E(Y|A=a,M,C) - m(a,a^*,C;\beta)\}\frac{I(A=a^*)}{f(A=a^*|C)}\right],$$

where $d_{a,a^*}(C)$ is an arbitrary index function. Because the mediation model is a union model corresponding to all possible choices a and a^*, all influence functions under the mediation model for arbitrary a and a^* are given by

$$\sum_a \sum_{a^*} d_{a,a^*}(C)\left[\{Y - E(Y|A=a,M,C)\}\frac{I(A=a)}{f(A=a|C)}\frac{f(M|A=a^*,C)}{f(M|A=a,C)} + \{E(Y|A=a,M,C) - m(a,a^*,C;\beta)\}\frac{I(A=a^*)}{f(A=a^*|C)}\right].$$

It follows from the above that for binary A taking the values 0 and 1, all influence functions

for β are of the form

$$
\begin{aligned}
& d_{0,0}(C)\frac{I(A=0)}{f(A=0|C)}\{Y-m(0,0,C;\beta)\} \\
& + d_{0,1}(C)\left[\frac{I(A=0)}{f(A=0|C)}\frac{f(M|A=1,C)}{f(M|A=0,C)}\{Y-E(Y|A=0,M,C)\}\right. \\
& \left.+\frac{I(A=1)}{f(A=1|C)}\{E(Y|A=0,M,C)-m(0,1,C;\beta)\}\right] \\
& + d_{1,0}(C)\left[\frac{I(A=1)}{f(A=1|C)}\frac{f(M|A=0,C)}{f(M|A=1,C)}\{Y-E(Y|A=1,M,C)\}\right. \\
& \left.+\frac{I(A=0)}{f(A=0|C)}\{E(Y|A=1,M,C)-m(1,0,C;\beta)\}\right] \\
& + d_{1,1}(C)\frac{I(A=1)}{f(A=1|C)}\{Y-m(1,1,C;\beta)\}\,.
\end{aligned}
$$

Writing

$$d_{a,a^*}(C) = f(A=a^*|C)d^*_{a,a^*}(C),$$

we obtain,

$$
\begin{aligned}
& I(A=0)d^*_{0,0}(C)\{Y-m(0,0,C;\beta)\} \\
& + I(A=0)d^*_{1,0}(C)\{E(Y|A=1,M,C)-m(1,0,C;\beta)\} \\
& + I(A=1)d^*_{0,1}(C)\{E(Y|A=0,M,C)-m(0,1,C;\beta)\} \\
& + I(A=1)d^*_{1,1}(C)\{Y-m(1,1,C;\beta)\} \\
& + \frac{I(A=0)f(A=1|C)}{f(A=0|C)}\frac{f(M|A=1,C)}{f(M|A=0,C)}d^*_{1,0}(C)\{Y-E(Y|A=0,M,C)\} \\
& + \frac{I(A=1)f(A=0|C)}{f(A=1|C)}\frac{f(M|A=0,C)}{f(M|A=1,C)}d^*_{0,1}(C)\{Y-E(Y|A=1,M,C)\}\,.
\end{aligned}
$$

In the proposed procedure, $d^*_{a,a^*}(C)$ is taken to be the vector of covariates appearing in the outcome model, including 1 for the intercept. The impact of the second step of the proposed algorithm is then to set the sample average of the last 2 lines in the above expression equal to zero. The parameters of interest can now be estimated by setting the sample average of the

remaining components in this expression to zero and solving for β. It is easily seen that this amounts to the proposed imputation procedure in lines 3-5 of the proposed algorithm.

Demonstration of double robustness

In this section, we confirm that the proposed estimation strategy yields a doubly robust: estimator. The development is again similar to that in Tchetgen and Schpitser (2011). Suppose first that the outcome model for Y is correctly specified. Then

$$
\begin{aligned}
& E\left[\{Y - E(Y|A=a,M,C)\}\frac{I(A=a)}{f(A=a|C)}\frac{f(M|A=a^*,C)}{f(M|A=a,C)}\right. \\
& \left. + \{E(Y|A=a,M,C) - m(a,a^*,C;\beta)\}\frac{I(A=a^*)}{f(A=a^*|C)}|C\right] \\
& = E\left[\{E(Y(a,M(a^*))|A=a,M(a^*),C) - m(a,a^*,C;\beta)\}\frac{I(A=a^*)}{f(A=a^*|C)}|C\right] \\
& = E\left[\{E(Y(a,M(a^*))|A=a^*,M,C) - m(a,a^*,C;\beta)\}\frac{I(A=a^*)}{f(A=a^*|C)}|C\right] \\
& = E\left[\{E(Y(a,M(a^*))|A=a^*,C) - m(a,a^*,C;\beta)\}\frac{I(A=a^*)}{f(A=a^*|C)}|C\right] \\
& = E\left[\{E(Y(a,M(a^*))|C) - m(a,a^*,C;\beta)\}\frac{I(A=a^*)}{f(A=a^*|C)}|C\right] = 0,
\end{aligned}
$$

which confirms the unbiasedness of the estimating functions. Suppose next that the models for M and A are correctly specified, then with $\tilde{E}(Y|A=a,M,C)$ denoting the possibly misspeci-

fied model,

$$
\begin{aligned}
&E\left[\left\{Y-\tilde{E}(Y|A=a,M,C)\right\}\frac{I(A=a)}{f(A=a|C)}\frac{f(M|A=a^*,C)}{f(M|A=a,C)}\right.\\
&\left.+\left\{\tilde{E}(Y|A=a,M,C)-m(a,a^*,C;\beta)\right\}\frac{I(A=a^*)}{f(A=a^*|C)}|C\right]\\
&=E\left[\left\{E(Y|A=a,M,C)-\tilde{E}(Y|A=a,M,C)\right\}\frac{I(A=a)}{f(A=a|C)}\frac{f(M|A=a^*,C)}{f(M|A=a,C)}\right.\\
&\left.+\left\{\tilde{E}(Y|A=a,M,C)-m(a,a^*,C;\beta)\right\}\frac{I(A=a^*)}{f(A=a^*|C)}|C\right]\\
&=\int\left\{E(Y(a,M(a^*))|C)-\int\tilde{E}(Y|A=a,M,C)f(M|A=a^*,C)dM\right\}\frac{I(A=a)}{f(A=a|C)}dA\\
&+E\left[\left\{\tilde{E}(Y|A=a,M,C)-m(a,a^*,C;\beta)\right\}\frac{I(A=a^*)}{f(A=a^*|C)}|C\right]\\
&=E(Y(a,M(a^*))|C)-\int\tilde{E}(Y|A=a,M,C)f(M|A=a^*,C)dM\\
&+\int\tilde{E}(Y|A=a,M,C)f(M|A=a^*,C)dM-m(a,a^*,C;\beta)=0,
\end{aligned}
$$

which shows that the estimating functions maintain their unbiasedness.

R-code for the data analysis

In the first step of the analysis we fit models for the outcome, mediator and exposure. For the outcome and exposure we defined logistic regression models:

```
my <- glm(Y ~ A + M + C , family = binomial)
ma <- glm(A ~ M + C , family = binomial)
```

and a linear regression model for the mediator:

```
mm <- lm(M  ~ A + C)
```

The **IPW-estimator** of the natural direct and indirect effect odds ratio is obtained from a model:

```
model_final = glm(Y ~ A + Astar + C, family=binomial,data=newDAT,
              weights=W)
```

where the dataset `newDAT` is constructed by repeating each observations in the original data set twice with an additional variable `Astar`, which is equal to the original exposure for the first replication and equal to the opposite of the actual exposure for the second replication. The weights `W` are obtained as follows:

```
W = as.numeric(dnorm(M,M_ASTAR,summary(mm)$sigma)/dnorm(M,fitted(mm),
      summary(mm)$sigma))
```

where `M` is the observed mediator value and `M_Astar` is a prediction from the model for the mediator with `A=Astar` and calculated as:

```
M_ASTAR=predict.lm(mm,newDAT)
```

The standard error, can be obtained via the bootstrap or software for generalized estimating equations by fitting the final model with the `geeglm()` procedure (from the `geepack` package) with an independence working correlation instead of the glm() procedure.

The **mean imputation estimator** is obtained by constructing a new outcome variable Y_{new} which is equal to the observed outcome if $A = a^*$ and equal to $E(Y|A = a^*, M, C)$ if $A \neq a^*$:

```
Ynew=ifelse(newDAT$A==newDAT$Astar,newDAT$Y,predict.glm(my,newDAT,
      type=c("response")))
```

This outcome is then used in next model

```
model_final.imp = glm(Ynew ~ A + Astar + C, family=binomial,
                      data=newDAT)
```

The standard error, can be obtained via the bootstrap

The **doubly robust estimator** uses a weighted model for the outcome:

```
myDR <- glm(Y ~ A + M + C , family = binomial,weights=WDR)
```

with corresponding weights WDR equal to:

```
 WDR = as.numeric((A*pA0*f0)/(pA1*f1) + ((1-A)*pA1*f1)/(pA0*f0))
```

where

```
pA1=fitted.values(model_A)
pA0 = 1-pA1


f1=as.numeric(dnorm(M,newdat$M_ASTAR[newdat$Astar==1],
     summary(mm)$sigma))
f0=as.numeric(dnorm(M,newdat$M_ASTAR[newdat$Astar==0],
      summary(mm)$sigma))
```

Based on the weighted model for the outcome, we construct a new outcome variable Y_{newDR} which is equal to the observed outcome if $A = a^*$ and equal to $E(Y|A = a^*, M, C)$ if $A \neq a^*$:

```
YnewDR=ifelse(newDAT$A==newDAT$Astar,newDAT$Y,
       predict.glm(myDR,newDAT,type=c("response")))
```

The new outcome is then used in the model

```
model_final.DR = glm(YnewDR ~ A + Astar + C, family=binomial,
       data=newDAT)
```

The standard errors of the three estimators can be, can be obtained via the bootstrap. Before we can perform a bootstrap we need to define a function which embeds the analysis.

```
anal1 <- function(data,index){
  dat=data[index,]


#fit working models
  model_Y =glm(Y ~ A + M + C,family=binomial,data=dat)
```

```
  mm <- lm(M  ~ A + C,data=dat)
  model_A =glm(A ~ C=binomial,data=dat)
  newdat=rbind(dat,dat)
  newdat$Astar=c(rep(1,dim(dat)[1]),rep(0,dim(dat)[1]))
#!! rename Astar to A
  newdat$M_ASTAR=predict.lm(mm,newdat)
  newdat$W = as.numeric(dnorm(newdat$M,newdat$M_ASTAR,summary(mm)$sigma)/
   dnorm(newdat$M,fitted(mm),summary(mm)$sigma))
  pA1=fitted.values(model_A)
  pA0 = 1-pA1
  f1=as.numeric(dnorm(dat$M,newdat$M_ASTAR[newdat$Astar==1],
  summary(mm)$sigma))
  f0=as.numeric(dnorm(dat$M,newdat$M_ASTAR[newdat$Astar==0],
  summary(mm)$sigma))
  WDR = as.numeric((dat$A*pA0*f0)/(pA1*f1) + ((1-dat$A)*pA1*f1)/
  (pA0*f0))
  model_YDR =glm(Y ~ A + M + C,weights=WDR,family=binomial,data=dat)

  #do the imputations
  newdat$YstarnoW=predict.glm(model_Y,newdata,type=c("response"))
  newdat$YnewnoW=ifelse(newdat$A==newdat$Astar,newdat$Y,
  newdat$YstarnoW)
  newdat$YstarDR=predict.glm(model_YDR,newdata,type=c("response"))
  newdat$YnewDR=ifelse(newdat$A==newdat$Astar,newdat$Y,
  newdat$YstarDR)

#fit the final models
```

```
model_final = glm(Y ~ A + Astar + C ,family=binomial,
data=newdat, weights=W)
model_finalimp = glm(YnewnoW ~ A + Astar + C ,family=binomial,
data=newdat)
model_finalDR = glm(YnewDR ~ A + Astar + C,family=binomial,
data=newdat)

#save the estimates of intrest
c(model_Y$coeff[2], model_Y$coeff[3]*mm$coeff[2],
\diamondmodel_Ytot$coeff[2]-model_Y$coeff[2],
model_final$coef[2], model_final$coef[3],
model_finalimp$coef[3],model_finalimp$coef[2],
model_finalDR$coef[3],model_finalDR$coef[2])
}
```

The bootstrap itself is done as follows:

```
library(boot)
b1 <- boot(anal1,data = dataset,R=1000)
```

Estimator	Level	OR	95% CI	P-value
Dampness and mold		1.36	1.10 to 1.67	0.0040
Perception control		1.31	1.19 to 1.43	< 0.0001
Age		1.02	1.01 to 1.03	< 0.0001
Sex	Women	1.63	1.34 to 1.99	< 0.0001
Crowding	> 1 per room	1		
	< 0.5	0.95	0.74 to 1.22	0.68
	0.51-1	1.42	1.02 to 1.99	0.040
Light	enough	1.29	1.04 to 1.59	0.017
City	Angers	1		
	Bonn	0.66	0.34 to 1.25	0.21
	Bratislava	1.92	1.17 to 3.21	0.012
	Budapest	2.41	1.50 to 3.98	0.0004
	Ferreira	4.08	2.62 to 6.58	< 0.0001
	Forli	1.00	0.61 to 1.69	0.99
	Geneva	1.01	0.54 to 1.85	0.97
	Vilnius	1.64	1.03 to 2.67	0.041
Marital status	married	1		
	separated	1.34	1.02 to 1.74	0.032
	single	0.91	0.67 to 1.23	0.55
Education	primary	1		
	secondary	0.86	0.67 to 1.11	0.25
	higher	0.49	0.35 to 0.68	< 0.0001
Env. tobacco smoke		1.54	1.26 to 1.88	< 0.0001

Table 5.10: Results from logistic outcome regression model.

Estimator	Level	OR	95% CI	P-value
Dampness and mold		1.22	1.15 to 1.28	< 0.0001
Age		0.99	0.99 to 1.00	< 0.0001
Crowding	> 1 per room	1		
	< 0.5	1.33	1.25 to 1.41	< 0.0001
	0.51-1	1.71	1.56 to 1.87	< 0.0001
Light	enough	1.11	1.04 to 1.17	0.0007
City	Angers	1		
	Bonn	0.73	0.65 to 0.81	< 0.0001
	Bratislava	1.05	0.94 to 1.17	0.42
	Budapest	0.60	0.53 to 0.67	< 0.0001
	Ferreira	1.10	0.98 to 1.24	0.11
	Forli	0.93	0.83 to 1.03	0.17
	Geneva	0.97	0.86 to 1.10	0.64
	Vilnius	1.14	1.02 to 1.26	0.019
Marital status	married	1		
	separated	1.08	1.00 to 1.17	0.055
	single	1.37	1.27 to 1.47	< 0.0001
Education	primary	1		
	secondary	1.10	1.02 to 1.18	0.016
	higher	1.11	1.02 to 1.20	0.016
Smoke	> 15 cigarettes	1		
	No	1.02	0.94 to 1.10	0.65
	Occasionally	1.00	0.92 to 1.08	0.92
	5-15 cigarettes	1.13	1.04 to 1.23	0.0043
Home type		1.13	1.05 to 1.22	0.0008

Table 5.11: Results from logistic mediator model.

Estimator	**Level**	**OR**	**95% CI**	**P-value**
Age		0.99	0.99 to 1.00	0.0002
Crowding	>1 per room	1		
	<0.5	1.66	1.43 to 1.93	<0.0001
	0.51-1	3.18	2.58 to 3.94	<0.0001
Light		1.28	1.12 to 1.47	0.0003
City	Angers	1		
	Bonn	0.54	0.42 to 0.70	<0.0001
	Bratislava	0.26	0.20 to 0.35	<0.0001
	Budapest	0.32	0.25 to 0.43	<0.0001
	Ferreira	2.93	2.11 to 4.07	<0.0001
	Forli	0.96	0.76 to 1.23	0.76
	Geneva	0.32	0.24 to 0.42	<0.0001
	Vilnius	0.41	0.31 to 0.52	<0.0001
Marital status	married	1		
	separated	1.06	0.88 to 1.27	0.56
	single	0.78	0.66 to 0.93	0.0059
Home type		1.82	1.53 to 2.17	<0.0001
Home size	$150m^2$	1		
	$0m^2$-$49m^2$	0.82	0.69 to 0.98	0.029
	$50m^2$ - $99m^2$	0.61	0.49 to 0.76	<0.0001
	$100m^2$-$149m^2$	0.53	0.38 to 0.72	<0.0001
Ventilation	free	1		
	No	0.71	0.60 to 0.84	0.0001
	Forced	0.88	0.76 to 1.01	0.062
Heating		1.71	1.40 to 2.08	<0.0001

Table 5.12: Results from logistic exposure model.

6

General Conclusion

During my four years of research I have tried to find an answer to the question if patients die *with* or *from* infection by:

1. Developing a new method which acknowledges the complex feedback relationship between severity of illness and infection and additionally takes into account the fact that patients who are discharged from the ICU are not comparable with those still in the ICU at that time.

2. Applying the method in the ICU-literature with the goal to convince physicians that the complex nature of ICU-data requires innovative statistical methods for analysis.

3. Bringing our message to the people (society) outside academia

In the final chapter of this thesis I look back at the challenges, achievements and some difficulties I have faced. I also give a general overview how the insights from the *model selection* and *mediation* part of my theses relate to the developed methodology for the analysis of ICU-data and finally I also point out some possible directions for future research and ways to further improve the estimation of the attributable mortaltiy of nosocomial infection.

It is not always about *being right*...

To me, an important achievement of this thesis is that, at some level, we have been able to convince a number of medical experts that a *more accurate* answer to the question if patients die *with* or *from* infection, requires acknowledging the dynamic relationship between severity of illness and infection using a causal analysis. We not only addressed the problem from a methodological point of view but directly introduced and applied our newly developed method into the ICU-literature. The key factor in the whole process was the close collaboration between statisticians and physicians with a strong belief in statistics and a lot of bedside experience in the ICU.

An important thing I have learned is that it is not always about *being right (gelijk hebben)* but actually a matter of *proving right (gelijk krijgen)*. In my opinion, statistical research is (too) often about *being right*. With mathematical proofs and simulations one can show that a newly developed method performs better than e.g. a standard approach. Once this is achieved, there is a certain chance that this will result in a publication in the statistical literature. Some theoretical developments can have a major impact and are necessary steps in the development of other methods. But some methods, like ours, are explicitly developed to answer practical (medical) questions. For that reason, it was important that after acceptance in the statistical literature (a matter of *being right*) we had a strategy to convince ICU-experts (a matter of *proving right*) that causal analysis is required to answer questions about the attributable mortality of nosocomial

infection. This was a very difficult but nevertheless important and very interesting step because if one is not able to convince researchers of the need for state-of-the-art statistical methodology, then the hard work invested in the development of the methods is somewhat lost.

Skepticism about causal inference.

The steps that were taken from the development of the methodology to the successful application in the ICU-literature were not without difficulties. The first step in this process was a letter to the editor of *Intensive Care Medicine* (Bekaert et al., 2010a) where we joined Wolkewitz et al. (2009) that statistical analysis should respect the time dependent nature of exposures, such as infection, but emphasized that the analysis of time-varying exposures also calls for specialized methods of confounder control which could be found in the field of causal inference. In their reply (Wolkewitz et al., 2010) the authors gave the message that, in an ICU-setting, a causal analysis is not appropriate because of the need for additional assumptions. In fact, also at a more general level, causal analyses were often criticized for the assumptions which they explicate. Many researchers then use more standard associational analysis instead, but nevertheless conclude with a causal interpretation of the results. Not only are such analyses hiding the necessary causal assumptions, the assumptions required for a causal interpretation to be justified are also typically much stronger. In a causal analysis which appropriately adjusts for confounding by including more detailed information about the *daily health condition of the patients*, the typical no unmeasured confounders assumption (which is also made by standard analysis) becomes more likely to be satisfied. Note furthermore that, an association is of little account when no causal interpretation can be given, because the relevant scientific question is a causal one: *do patients die from (i.e., due to) or with infection*?

Personally I have the feeling that to many people (statisticians) who are not familiar with causal inference think that it is only about the development and use of very complex methods involving

weights and other fancy estimation procedures. In the ICU-setting, a causal analysis is mainly about confounding adjustment (see **chapter 5** for another focus) which in simple settings can be done using standard regression techniques. In more complex settings (like ours) standard regression approaches fail to appropriately account for confounding so we need to use more complex (non-standard) methods. A major difference between a standard analysis and a causal analysis is that researchers in causal inference explicitly think and communicate about the assumptions they made. Every statistical analysis which aims for a causal interpretation assumes e.g. that all confounders for the *exposure-outcome relationship* are measured or that e.g the treatment or exposure is randomized so that confounding can be ignored. This is no more than the unmeasured confounders assumption necessary to link the observed data to the counterfactuals (see **chapter 1** for a detailed description). A second important assumption in causal inference is the *consistency assumption* which, in our nosocomial infection setting states that, if we were able to do an intervention in which it was possible to prevent infection for all, this would be a non-invasive one in which it would have the same consequences as if the patients would naturally prevent infection. Or, if we could infect patients, this would have the same effect on their outcome as they would naturally acquire infection.

Actually, the most important phase in the analysis is the one before the analysis of the data, where we think about the *data generating mechanism* and how plausibly the assumptions are met. In this first step, it is important to distinguish *confounders* from e.g. *mediators*. Expert knowledge about the scientific topic is very important because the rest of the analysis relies on this step. Therefore a close collaboration between statisticians and physicians with a lot of expertise and bedside experience with nosocomial infections and the link with underlying severity of illness is very important and played an important role in this thesis.

Causal inference in the ICU-literature

At the end of 2009 a review from Melsen et al. (2009) concluded that their remains much controversy about the estimation of the attributable mortality of VAP. In an editorial (Muscedere, 2009) on the matter it was stated that *"Because observational data in regard to the attributable mortality of VAP are all that we will ever have, studies of more methodological rigor are required if we are going to answer this important question."* Keeping this in mind, we have tried to convince a number of ICU-experts about the need of more sophisticated statistical methods and published a study with *more methodological rigor* in the world's leading intensive care unit journal. During that process, we found a good balance between the methodological and medical relevant messages.

Personally I am very glad to read the conclusion of the most recent review on the attributable mortality of VAP (Timsit et al., 2011) which stated that: *"Classical statistical analyses of VAP attributable mortality have ignored the aforementioned issues of time-dependent confounding and empirical results are therefore highly controversial, with several studies reporting estimates for mortality ranging from neutral to severely harmful. To address these problems, many innovative statistical approaches have been proposed that would shed new light on our understanding of the role of VAP in ICU mortality and maximal efforts should be made for promoting their uses among statisticians and clinicians and their appearance in medical journals."* A goal of this thesis was to convince ICU-experts that the complex nature of ICU-data requires innovative statistical methods for analysis. I am glad that I can end my four years of research with a high quality medical publication and above message published in the ICU-literature. On the other hand it would not be fair to claim that we already achieved our goal because the critical reader will notice that Prof. Timsit is, as a VAP and ICU-expert, also one of the coauthors of our medical paper. An important strategy in the process from *being right* to *proving right* is to convince the *right* people so that they can further spread our message.

Bringing our message to the people.

dS De Standaard *voor abonnees*

NIEUWS INFOTHEEK E-KRANT ARCHIEF+ DIENSTEN IPAD/IPHONE DELUXE

voor abonnees » infotheek » Binnenland

Ziekenhuisinfecties eisen 2.600 doden per jaar

maandag 02 februari 2009 Bron: belga sdg

AANRADEN 1

Ziekenhuisinfecties zorgen ervoor dat in België elk jaar 2.600 patiënten voortijdig sterven tijdens hun verblijf in het ziekenhuis. Dat blijkt uit cijfers van het Federaal Kenniscentrum voor de Gezondheidszorg (KCE).

belga 1 van 2

Ziekenhuisinfecties vormen de meest voorkomende complicatie bij gehospitaliseerde patiënten. Het KCE wist vorig jaar al te melden dat 6 procent van alle ziekenhuispatiënten kampt met een ziekenhuisinfectie. Naar schatting 2.600 van hen overlijden jaarlijks aan de aandoening, berekende het Kenniscentrum nu.

Daarnaast zorgt een ziekenhuisinfectie ervoor dat het verblijf in het hospitaal met gemiddeld een week wordt verlengd. Dat kost de maatschappij jaarlijks 384 miljoen euro.

...

February 2009

dS De Standaard *voor abonnees*

NIEUWS INFOTHEEK E-KRANT ARCHIEF+ DIENSTEN IPAD/IPHONE DELUXE

voor abonnees » infotheek » Binnenland

'Overlijdens door ziekenhuisbacterie ernstig overschat'

dinsdag 23 augustus 2011, 10u54 Bron: Belga dar

AANRADEN 30

Een UGent-onderzoeksteam van statistici en artsen, werkzaam binnen intensieve zorgen en onder leiding van professor Dominique Benoit, heeft een methode ontwikkeld om overlijdens ten gevolge van de ziekenhuisbacterie correcter in kaart te brengen. Dat aantal overlijdens wordt ernstig overschat, zo menen ze.

'De vraag is of een patiënt sterft ten gevolge van of eerder met de infectie', zegt Benoit. 'Binnen de vakliteratuur bestaat heel wat onduidelijkheid door de complexe relatie tussen de infectie en het onderliggend ziektebeeld van de patiënt.'

De nieuwe methode is de eerste die daarmee rekening houdt. Ze werd toegepast om de impact van de ernstigste infecties binnen een groep van 4.479 mechanisch beademende patiënten uit elf Franse eenheden voor intensieve zorgen te onderzoeken.

'Gebaseerd op deze data stierf volgens onze berekening de meerderheid van de patiënten met de infectie en zouden er per jaar maar 7 patiënten gestorven zijn door de infectie. Door het ontbreken van kwaliteitsvolle data op Belgisch niveau kunnen we moeilijk uitspraken doen, maar duidelijk is dat het aantal van naar schatting 2.600 Belgen die jaarlijks

...

August 2011

In 2009, it was estimated (Vrijens et al., 2009) that nosocomial infections in Belgium have an *excess mortality* of 2625 deaths per year and *excess costs* for the healthcare payer of nearly 400 € million per year. These numbers are estimated from observed data without taking into account the dynamic (time-dependent) relationship between severity of illness and infection. Without appropriate confounding control one is estimating associations instead of causal relationships. Nevertheless, the message was brought as a causal one and got a lot of media attention (see e.g. a press release on February 2009 in 'De Standaard').

After the publication of our medical paper a next and final step was the communication of our message outside academia because our results suggested that the numbers from the earlier report were overestimating the attributable mortality of nosocomial infection. Although we could not give a more accurate estimate for the earlier published numbers we were able to bring our main

message that there is a difference in patients dying *with* or *from* infections depending on their underlying severity of illness evolving over time. (see e.g. a press release on August 2011 in 'De Standaard').

After four years of working on the topic, I somewhat have the feeling that my mission is completed: we started with an important medical question, we developed a new statistical methodology, introduced it in the ICU-literature to answer the question and brought our main message outside academia.

Combining the insights from all methodological developments

In this thesis we worked on three methodological topics from the field of causal inference:

1. In **chapter 2** we developed a method to estimate the causal effect of a time varying exposure in the presence of competing risks. Because our focus was mainly on the cumulative incidence function (CIF) we used a marginal structural model approach to model the subdistribution hazard (instead of modeling the cause specific hazard).

2. In **chapter 4** we developed a model selection procedure which directly targets the quality of the causal exposure effect estimator. Additionally we studied the impact of model misspecification the so-called G-estimators and inverse probability weighted (IPW) estimators of a causal effect.

3. In **chapter 5** we moved the focus to the estimation of natural direct and indirect effects and developed three estimators: a simple IPW-estimator, a doubly robust version and an imputation based estimators which doesn't involve weighting.

The method developed in **chapter 2** and applied in **chapter 3** is a two step approach in which the first step involves a model for infection conditional on the time varying severity of illness indicators. A stepwise selection procedure was used to select the severity of illness indicators

which were associated with infection. Based on this model, weights were calculated and used to fit a marginal strucutural model for the subdistribution hazard. In a next step the cumulative incidence function (CIF) was calculated and the attributable mortality, which was here the *effect estimand of interest*, was then calculated as the population attributable fraction. In fact, the best way to fit the models for the weights is unclear and was the main motivation to developed a model selection procedure which focuses on the quality of the causal estimate of interest. In **chapter 4** of this thesis we explored the ideas in a simple setting but it would theoretically possible to directly target the quality of the attributable mortality in terms of the mean squared error (MSE). The confounder space would consist of all possible time varying severity of illness indicators which can be possibly included in the model for the weights. In every orbit, the covariate is then selected which minimizes the MSE of the estimated population attributable fraction. Because we are not directly modeling the quantity of interest, several steps are involved in the estimation of it, and need to take into account in the calculation of the variance and the bias of our quantity of interest:

- The population attributable fraction is calculated from the cumulative incidence function;
- The cumulative incidence function is calculated from the subdistribution hazard under a marginal structural model;
- The weights used in the marginal structural model are calculated based on a model for the exposure;

In the calculation of the variance of the attributable mortality we took into account all these steps by using the theory of estimating equations. The major challenge is that the methodology and the calculation of the MSE and FIC has only been developed so far for maximum likelihood estimators. This means that additional theoretical work is needed to extend the estimation of the FIC to a setting based on estimating equations because it will not be feasible to translate the whole estimation problem into a likelihood based approach (Robins et al., 1999). From a practical point of view I see several potential problems. The combination of the orbit selection

algorithm with the algorithm to extend the dataset with all potential infection regimes will be computationally very intensive.

The estimator of the causal effect of infection is called an IPW-estimator because it involves weights. When those weights are highly variable those estimators may perform very badly and can be very inefficient. The strategy we used was to stabilize the weights via the use of semi parametric theory. This resulted in a more efficient estimator. Another possibility is the development of a doubly robust estimator, as in **chapter 5**, for the causal effect of a time varying exposure in the presence of competing risks. Such estimator would be protected against some model misspecification in one of both working models (for the exposure or the outcome). Using the property of a doubly robust estimator we can then truncate extreme weights and obtain more efficient estimators of the effect of interest without worrying about the misspecification of the model for the weights. As we saw in **chapter 5** the imputation estimator of the natural direct and indirect effect, which was obtained after extreme truncation of the weights (setting all weights to 1) performed very well. The development of such estimators in the context of the more could be an additional topic for future research.

Further research

Future collaborations between statisticians and physicians

One conclusion of this thesis is that a close collaboration between statisticians and physicians can be beneficial for both parties. We were able to introduce our methodological ideas from the statistical into the ICU-literature and they were able to bring an important medical message. The success of state-of-the art statistical methods for confounding adjustment completely relies on the quality of the data. Current surveillance systems have important limitations such as inaccurate, incomplete and partly retrospective data, resulting in labour-intensive analysis and a limited integration of the information regarding the matching of the nosocomial infection with

the associated microbiology, given antibiotics and underlying severity of illness. The future lies in the detailed translation of the health condition, in terms of physio-pathology and therapeutic treatment, of the patient into high quality, longitudinal data used as a source for appropriate statistical techniques.

One example is the COSARA project leaded by prof. dr. Johan Decruyenaere from the Ghent University hospital. COSARA stands for **Co**mputer-based **S**urveillance and **A**lerting of Nosocomial Infections, Antimicrobial **R**esistance and **A**ntibiotic Consumption in the Intensive Care Unit. With the use of state-of-the-art ICT, the system completely automates the surveillance of nosocomial infections and is able to give alerts if alarming trends occur in the incidence of nocosomial infections, in the incidence of specific multi-resistant microorganisms or in the use of antimicrobial drugs. Besides giving the alert itself, it is able to generate automatically all information needed to interpret an alert easier. The cornerstone of COSARA is the *ICU Patient Infectious Agent* (IPIA), responsible for collecting on a daily basis all necessary information on the patient level. The IPIA is completely integrated in the already high-advanced ICU computerization. To ensure the quality of the data by daily compliance of the intensivist, IPIA will offers the physician important decision making support during his/her daily observations. The data from such innovative surveillance systems are extremely useful and make it possible to obtain further insights into the complex relationship of infection and underlying severity of illness.

Our method is not only useful to estimate the attributable mortality of nosocomial infections. It can be applied to every time dependent exposure (e.g. effect of noninvasive ventilation (NIV), DNR-codes) in the presence of complex feedback relationships between the exposure and time dependent confounders (severity of illness) in the presence of competing risks (ICU-setting with discharge as the competing event for ICU mortality). We mainly worked on the attributable mortality but other very interesting directions are to estimate attributable length of stay (LOS) in the ICU together with the attributable costs.

Extending the existing methodology to the estimation of the attributable mortality of VAP

Improvements can be done from two point of views. From a methodological point of view a doubly robust estimator for the causal effect of time varying exposures in the presence of competing risks could be derived and implemented in the software. An advantage of such estimators is that they are, at some level, protected against model misspecification of one of the working models for infection or mortality (see **chapter 5** for an application in a more simple setting of direct and indirect effects). One major improvement is needed in terms of the computational efficiency. In our analysis, the marginal structural model for the subdistribution hazard is fitted using a pooled logistic regression model in which the data is structured as one-line-per-patient-day. Because every patient contributes several potential infection regimes some data extension is needed and resulted for our example of 4479 patients in a final dataset of more than 2.5 million lines. Analysis of more patients over a longer follow up period would exceed the memory boundaries of R. One possible solution lies in the use of multiple imputation. For those who are discharged we know that they are compatible with several hypothetical infection regimes. Instead of extending the data with all possibilities we may randomly impute one possible infection regime. By doing so, we construct a dataset in which every patient is observed until death (the event of interest) or the end of follow up (for those who were discharged). If we repeat the analysis several times we can summarize the estimates by simply averaging all obtained estimates. In the calculation of the variance the between and within imputation variability is taken into account as in multiple imputation for missing data (Rubin, 1987).

In the discussion of **chapter 3** we give an overview of the limitations of the existing method (and a lot of other standard approaches). Marginal structural models model the causal effect of a time dependent exposure on the distribution of treatment-specific counterfactual outcomes of interest. It is a marginal model and the adjustment for time varying severity of illness indicators is done by using weights which are obtained from the first step of the analysis. There is a

strong belief that the infection effect depends on the evolution of severity of illness or antibiotic treatment. These relationships can be modeled by statistical interactions. However marginal structural models do not allow for interactions with time dependent covariates. If one is e.g. interested in the impact of nosocomial infections caused by different germs it is not appropriate to divide the population in subgroups defined by the type of germ because infection caused by a given germ is not a random process. It is e.g. possible that the type of infection is related to the underlying health status of the patient. What we can do is adjust the weights by additionally including the probability of acquiring a germ specific infection. The same reasoning holds for other time dependent characteristics, like appropriate treatment or not.

Another drawback of marginal structural models is that they compare the risk of death between two hypothetical infection paths in which everybody vs. nobody acquires infection. Such effect measures can be difficult to interpret because scenarios in which everybody acquires infection (via a hypothetical intervention) are not realistic. The most relevant intervention is therefore one where infection would be prevented for all patients and thus our substantive interest lies in the infection-free path $\overline{a}(0)$. This is why our definition of attributable mortality compares the counterfactual CIF under this infection path can be with the observed CIF.

In view of above limitations the future maybe lies in another class of models (structural nested models) which model the impact of postponing the infection with a day within the infected patients. For instance, with $\overline{A}_t$ and $\overline{L}_t$ the history of infection and confounders respectively, we can postulate that, on the logit scale, the difference in mortality risk due to postponing infection with one day is constant over time:

$$\text{logit}P\{Y(\overline{a}_{t-1}=0,a_t=1)=1|\overline{A}_{t-1}=0,\overline{A}_t=1,\overline{L}_t\}-$$
$$\text{logit}P\{Y(\overline{a}_t=0,a_{t+1}=1)=1|\overline{A}_{t-1}=0,\overline{A}_t=1,\overline{L}_t\}=\beta$$

In the above model the odds of dying would change with $\exp(\beta)$ if infection was postponed with 1 day. If we are able to combine models of this form on different time points t from the time

to infection until the end of follow up, we can calculate what the risk of dying would have been if infection would have been prevented for infected patients. Because the model incorporates time dependent covariates it would be possible to model interaction between infection and these time dependent confounders (severity of illness). In addition, by quantifying the infection effect within the infected, they yield possibly more meaningful causal parameters.

Statistical inference for logistic structural models has so far not been developed for longitudinal data and can be a very interesting topic for future research. Recently, Martinussen et al. (2011) made some progress in developments based on additive hazard models.

Scientific output

Awards

- The Best Student Presentation Award at the 2nd IBS Channel Network Conference (2009)
- Student Conference Award of the 30th Annual Conference of the International Society for Clinical Biostatistics (2009)

Publications

Published

Depuydt, P., Vandijck, D., Bekaert, M., Decruyenaere, J., Blot, S., Vogelaers, D. and Benoit, D. (2008): "Determinants and impact of multidrug antibiotic resistance in pathogens causing ventilator-associated-pneumonia," *Critical Care*, 12(6) R142.

Bekaert, M., Vansteelandt, S. and Mertens, K. (2010): "Adjusting for time-varying confounding in the subdistribution analysis of a competing risk," *Lifetime Data analysis*, 16(1) 45-70.

Bekaert, M., Timsit, J.F., Vansteelandt, S., Depuydt, P., Vsin, A., Garrouste-Orgeas, M., Decruyenaere, J., Clec'h, C., Azoulay, E. and Benoit, D. (2011): "Attributable mortality of ventilator associated pneumonia : a reappraisal using causal analysis," *American Journal of Respiratory and Critical Care Medicine* doi: 10.1164/rccm.201105-0867OC.

Bekaert, M., Benoit, D., Decruyenaere, J. and Vansteelandt, S. (2010): "A note on statistical association and causality derived from epidemiological ICU data : reply," *Intensive Care Medicine*, 36(3) 550.

Vansteelandt, S., Bekaert, M. and Claeskens, G. (2011): "On model selection and model mis-

specification in causal inference," *Statistical Methods in Medical Research* doi: 10.1177/0962 280210387717.

Duytschaever, G., Huys, G., Bekaert, M., Boulanger, L., De Boeck, K. and Vandamme, P. (2012): "Cross-sectional and longitudinal comparison of the predominant fecal microbiota composition between a group of pediatric patients with cystic fibrosis and their healthy siblings," *Applied and Environmental Microbiology* doi: 10.1128/AEM.05933-11.

Submitted work

Two papers on the estimation of natural direct and indirect effect are submitted and currently under review. I was also involved as the statistician in the Appropricus Study. The results of this study are currently under review in the Journal of the American Medical Association (JAMA).

Presentation at conferences

The results of my research were presented at several national and international conferences.

National conferences

- IAP workshop in 2008 and 2010
- meeting of the Belgian Statistical Society (BSS) in 2008, 2010 and 2011
- IBS channel meeting in (2009)

International conferences

- Meeting of the Eastern North American Region (ENAR) of the International Biometrical Society in Texas, (2009)
- Meeting of the European Society of Intensive Care Medicine (ESICM) in Vienna (2009)

- The annual conference of the International Society for Clinical Biostatistics (ISCB) in Prague (2009) and Montpellier (2010)
- International Biometric Conference (IBC) in Florianopolis (2010)
- Workshop causal inference, Montreal (2011)

References

Andersen, P., O. Borgan, R. Gill, and K. N (1993): *Statistical models based on counting processes*, Springer Series in Statistics New York.

Andersen, P. K., S. Z, and S. Rosthj (2002): "Competing risks as a multi-state model," *Statistical Methods in Medical Research*, 11, 203–215.

Barnett, A. and N. Graves (2008): "Competing risks models and time-dependent covariates," *Critical Care*, 12, 143.

Baron, R. and D. Kenny (1986): "The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations," *J. Pers. Soc. Psychol.*, 51, 1173–1182.

Bekaert, M., D. Benoit, J. Decruyenaere, and S. Vansteelandt (2010a): "A note on statistical association and causality derived from epidemiological icu data reply," *Intensive Care Medicine*, 36, 550–550.

Bekaert, M., J.-F. Timsit, S. Vansteelandt, P. Depuydt, A. Vesin, M. Garrouste-Orgeas, J. Decruyenaere, C. Clec'h, E. Azoulay, and D. Benoit (2011): "Attributable mortality of ventilator associated pneumonia: A reappraisal using causal analysis," *Am. J. Respir. Crit. Care Med.*, 201105–0867OC.

Bekaert, M., S. Vansteelandt, and K. Mertens (2010b): "Adjusting for time-varying confounding in the subdistribution analysis of a competing risk," *Lifetime Data Analysis*, 16, 45–70.

Bercault, N. and T. Boulain (2001): "Mortality rate attributable to ventilator-associated nosocomial pneumonia in an adult intensive care unit: a prospective case-control study," *Critical Care*, 29, 2303–2309.

Beyersmann, J., P. Gastmeier, H. Grundmann, S. Bärwolff, C. Geffers, M. Behnke, H. Rüden, and M. Schumacher (2006): "Use of multistate models to assess prolongation of intensive care unit stay due to nosocomial infection," *Infection Control and Hospital Epidemiology*, 27, 493–499.

Beyersmann, J. and M. Schumacher (2008): "Time-dependent covariates in the proportional subdistribution hazards model for competing risks," *Biostatistics*, 9, 765–776.

Beyersmann, J., M. Wolkewitz, and M. Schumacher (2008): "The impact of time-dependent bias in proportional hazards modelling," *Stat med*, 27, 6439–6454.

Bonten, M. J. M., M. H. Kollef, and J. B. Hall (2004): "Risk factors for ventilator-associated pneumonia: From epidemiology to patient management," *Clinical Infectious Diseases*, 38, 1141–1149.

Brookhart, M., S. Schneeweiss, K. Rothman, R. Glynn, J. Avorn, and T. Sturmer (2006): "Variable selection for propensity score models," *American Journal of Epidemiology*, 163, 1149–1156.

Brookhart, M. A. and M. J. van der Laan (2006): "A semiparametric model selection criterion with applications to the marginal structural model," *Computational Statistics & Data Analysis*, 50, 475–498.

Budtz-Jorgensen, E., N. Keiding, P. Grandjean, and P. Weihe (2007): "Confounder selection

in environmental epidemiology: Assessment of health effects of prenatal mercury exposure," *Annals of Epidemiology*, 17, 27–35.

Buenocavanillas, A., M. Delgadorodriguez, A. Lopezluque, S. Schaffinocano, and R. Galvezvargas (1994): "Influence of nosocomial infection on mortality-rate in an intensive-care unit," *Critical Care Medicine*, 22, 55–60.

Cao, W., A. Tsiatis, and M. Davidian (2009): "Improving efficiency and robustness of the doubly robust estimator for a population mean with incomplete data," *Biometrika*, 96, 723–734.

Carlet, J. (2001): "Dying from or with a nosocomial pneumonia in the intensive care unit?" *Critical Care Medicine*, 29, 2392–2394.

Chamberlain, G. (1987): "Asymptotic efficiency in estimation with conditional moment restrictions," *Journal of Econometrics*, 34, 305–334.

Claeskens, G., C. Croux, and J. Van Kerckhoven (2006): "Variable selection for logistic regression using a prediction focussed information criterion," *Biometrics*, 62, 972–979.

Claeskens, G. and N. L. Hjort (2003): "The focused information criterion," *Journal of the American Statistical Association*, 98, 900–916, with discussion and a rejoinder by the authors.

Claeskens, G. and N. L. Hjort (2008): *Model Selection and Model Averaging*, Cambridge: Cambridge University Press.

Connors, A. F., T. Speroff, N. V. Dawson, C. Thomas, F. Harrell, D. Wagner, N. Desbiens, L. Goldman, A. Wu, R. M. Califf, W. J. Fulkerson, H. Vidaillet, S. Broste, P. Bellamy, J. Lynn, , and W. A. Knaus (1996): "The effectiveness of right heart catheterization in the initial care of critically ill patients," *Journal of the American Medical Association*, 11, 889–897.

Crainiceanu, C., F. Dominici, and G. Parmigiani (2008): "Adjustment uncertainty in effect estimation," *Biometrika*, 95, 635–651.

D'Agostino, R., Jr. (1998): "Propensity score methods for bias reduction in the comparison treatment to a non-randomized control group," *Statistics in Medicine*, 17, 2265–2281.

Davis, K. (2006): "Ventilator-associated pneumonia: A review," *Journal of Intensive Care Medicine*, 21, 211–226.

De Luna, X., T. Richardson, and I. Waernbaum (2010): "Covariate selection for the non-parametric estimation of an average treatment effect," Technical report, Umea University.

Fagon, J., J. Chastre, A. Hance, P. Montravers, A. Novara, and C. Gibert (1993): "Nosocomial pneumonia in ventilated patient: a cohort study evaluating attributable mortality and hospital stay," *American Journal of Medicine*, 94, 281–299.

Fine, J. and R. Gray (1999): "A proportional hazards model for the subdistribution of a competing risk," *Journal of the American Statistical Association*, 94, 496–509.

Girou, E., F. Stephan, A. Novara, M. Safar, and J. Fagon (1998): "Risk factors and outcome of nosocomial infections: results of a matched case-control study of ICU patients," *American Journal Respiratory Critocam Care Medicine*, 157, 1151–1158.

Greenland, S. (1986): "Modeling and variable selection in epidemiologic analysis," *American Journal of Public Health*, 79, 340–349.

Greenland, S. (2003): "Quantifying biases in causal models: Classical confounding vs collider-stratification bias," *Epidemiology*, 14, 300–306.

Greenland, S. (2007): "Bayesian perspectives for epidemiological research. ii. regression analysis," *International Journal of Epidemiology*, 36, 195–202.

Greenland, S. (2008): "Variable selection versus shrinkage in the control of multiple confounders," *American Journal of Epidemiology*, 167, 523–529.

Greenland, S. and R. Neutra (1980): "Control of confounding in the assessment of medical technology," *International Journal of Epidemiology*, 9, 361–367.

Greenland, S., J. Pearl, and J. Robins (1999a): "Causal diagrams for epidemiological research." *Epidemiology*, 10, 37–48.

Greenland, S., J. Robins, and J. Pearl (1999b): "Confounding and collapsibility in causal inference," *Statistical Science*, 14, 29–46.

Hafeman, D. and S. Schwartz (2009): "Opening the black box: a motivation for the assessment of mediation," *International Journal of Epidemiology*, 38, 838–845.

Hahn, J. (2004): "Functional restriction and efficiency in causal inference," *The Review of Economics and Statistics*, 86, 73–76.

Haight, T., Y. Wang, M. van der Laan, and I. Tager (2010): "A cross-validation-deletion,-substitution,-addition model selection algorithm: Application to marginal structural models," *Computational Statistics and Data Analysis*, In press.

Hand, D. J. and V. Vinciotti (2003): "Local versus global models for classification problems: fitting models where it matters," *The American Statistician*, 57, 124–131.

Hernán, M. (2004): "A definition of causal effect for epidemiological research," *Journal of Epidemiology and Community Health*, 58, 265–271.

Hernán, M., B. Brumback, and J. Robins (2000): "Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men," *Epidemiology*, 11, 561–570.

Hernán, M., S. H. Diaz, and J. Robins (2004): "A structural approach to selection bias," *Epidemiology*, 15, 615–625.

Hernan, M., S. Hernandez-Diaz, M. Werler, and A. Mitchell (2002): "Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology," *American Journal of Epidemiology*, 155, 176–184.

Hernán, M. and J. Robins (2006): "Estimating causal effects from epidemiological data," *J Epidemiol Community Health*, 60, 578–586.

Hernán, M. A. (2010): "The hazards of hazard ratios," *Epidemiology*, 21, 13–15.

Hernán, M. A., A. Alonso, R. Logan, F. Grodstein, K. B. Michels, W. C. Willett, J. E. Manson, and J. M. Robins (2008): "Observational studies analyzed like randomized experiments an application to postmenopausal hormone therapy and coronary heart disease," *Epidemiology*, 19, 766–779.

Heyland, D., D. Cook, and L. Griffith (1999): "The attributable morbidity and mortality of ventilator-associated pneumonia in the critically ill patient," *American Journal Respiratory Critocam Care Medicine*, 159, 1249–1256.

Hjort, N. L. and G. Claeskens (2003): "Frequentist model average estimators," *Journal of the American Statistical Association*, 98, 879–899, with discussion and a rejoinder by the authors.

Hong, G. (2010): "Ratio of mediator probability weighting for estimating natural direct and indirect effects," in *Proceedings of the American Statistical Association, Biometrics Section*, Alexandria, VA: American Statistical Association, 2401–2415.

Hotchkiss, R. S. and S. Opal (2010): "Immunotherapy for sepsis - a new approach against an ancient foe," *New England Journal of Medicine*, 363, 87–89.

Imai, K., L. Keele, and D. Tingley (2010): "A General Approach to Causal Mediation Analysis," *Psychological Methods*, 15, 309–334.

Kalbfleisch, J. and R. Prentice (2002): *The statistical analysis of failure time data, 2nd edition*, Hoboken: John Wiley & Sons.

Kang, J. and J. Schafer (2008): "Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data," *Statistical Science*, 22, 523–539.

Keiding, N., M. Filiberti, S. Esbjerg, J. Robins, and N. Jacobsen (1999): "The graft versus

leukemia effect after bone marrow transplantation: A case study using structural nested failure time models," *Biometrics*, 55, 23–28.

Kim, P. W., T. M. Perl, E. F. Keelaghan, P. Langenberg, E. N. Perencevich, A. D. Harris, X. Y. Song, and M. C. Roghmann (2005): "Risk of mortality with a bloodstream infection is higher in the less severely ill at admission," *American Journal of Respiratory and Critical Care Medicine*, 171, 616–620.

Kollef, M. (2003): "The importance of appropriate initial antibiotic therapy for hospital-acquired infections." *Am J of Med*, 115, 582–584.

Koulenti, D., T. Lisboa, C. Brun-Buisson, W. Krueger, A. Macor, J. Sole-Violan, E. Diaz, A. Topeli, J. DeWaele, A. Carneiro, I. Martin-Loeches, A. Armaganidis, and J. Rello (2009): "Spectrum of practice in the diagnosis of nosocomial pneumonia in patients requiring mechanical ventilation in european intensive care units," *Critical Care Medicine*, 37, 2360–8.

Krueger, W. A., F. P. Lenhart, G. Neeser, G. Ruckdeschel, H. Schreckhase, H. J. Eissner, H. Forst, J. Eckart, K. Peter, and K. E. Unertl (2002): "Influence of combined intravenous and topical antibiotic prophylaxis on the incidence of infections, organ dysfunctions, and mortality in critically ill surgical patients: a prospective, stratified, randomized, double-blind, placebo-controlled clinical trial," *Am J Respir Crit Care Med*, 166, 1029–37.

Lange, T. and J. V. Hansen (2011): "Direct and Indirect Effects in a Survival Context," *Epidemiology*, 22, 575–581.

Lange, T., S. Vansteelandt, and M. Bekaert (2011): "A simple approach for estimating natural direct and indirect effects," Technical report, University of Copenhagen.

Latouche, A., R. Porcher, and S. Chevret (2005): "A note on including Time-dependent Covariate in Regression Model for Competing Risks Data," *Biometrical Journal*, 47, 807–814.

Legall, J. R., S. Lemeshow, and F. Saulnier (1993): "A new simplified acute physiology score (saps-ii) based on a european north-american multicenter study," *Jama-Journal of the American Medical Association*, 270, 2957–2963.

Leon, S., A. A. Tsiatis, and M. Davidian (2003): "Semiparametric estimation of treatment effect in a pretest-posttest study," *Biometrics*, 59, 1046–1055.

Little, R. (1985): "A note about models for selectivity bias," *Econometrica*, 53, 1469–1474.

MacKinnon, D. (2008): *An Introduction to Statistical Mediation Analysis*, New York: Lawrence Erlbaum Associates.

Maldonado, G. and S. Greenland (1993): "Simulation study of confounder-selection strategies," *American Journal of Epidemiology*, 138, 923–936.

Martinussen, T., S. Vansteelandt, M. Gerster, and J. v. B. Hjelmborg (2011): "Estimation of direct effects for survival data by using the aalen additive hazards model," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, no–no, URL `http://dx.doi.org/10.1111/j.1467-9868.2011.00782.x`.

Melsen, W. G., M. M. Rovers, and M. J. M. Bonten (2009): "Ventilator-associated pneumonia and mortality: A systematic review of observational studies," *Critical Care Medicine*, 37, 2709–2718.

Mickey, R. and S. Greenland (1989): "The impact of confounder selection criteria on effect estimation," *American Journal of Epidemiology*, 129, 125–137.

Mortimer, K., R. Neugebauer, M. van der Laan, and I. Tager (2005): "An application of model-fitting procedures for marginal structural models," *American Journal of Epidemiology*, 162, 382–388.

Muscedere, J. (2009): "Ventilator-associated pneumonia and mortality: The controversy continues," *Critical Care Medicine*, 37, 2845–2846.

Nguile-Makao, M., J. R. Zahar, A. Francais, A. Tabah, M. Garrouste-Orgeas, B. Allaouchiche, D. Goldgran-Toledano, E. Azoulay, C. Adrie, S. Jamali, C. Clec'h, B. Souweine, and J. F. Timsit (2010): "Attributable mortality of ventilator-associated pneumonia: respective impact of main characteristics at icu admission and vap onset using conditional logistic regression and multi-state models," *Intensive Care Medicine*, 36, 781–789.

Papazian, L., F. Bregeon, and X. Thirion (1996): "Effect of ventilator-associated pneumonia on mortality and morbidity," *American Journal Respiratory Critocam Care Medicine*, 154, 91–97.

Pearl, J. (1998): "Graphs, causality, and structural equation models," *Sociological Methods and Research*, 27, 226–284.

Pearl, J. (2001): "Direct and indirect effects," in *Proceedings of the Seventeenth Conference on Uncertainty and Artificial Intelligence*, San Francisco: Morgan Kaufmann, 411–420.

Pearl, J. (2009a): *Causality: Models, Reasoning and Inference*, Cambridge: Cambridge University Press.

Pearl, J. (2009b): "Remarks on the method of propensity score," *Statistics in Medicine*, 28, 1415–1416.

Pearl, J. (2010): "On a class of bias-amplifying covariates that endanger effect estimates," in P. Grunwald and P. Spirtes, eds., *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence*, AUAI, Corvallis, OR, 417–424.

Pearl, J. (2011): "The mediation formula: A guide to the assessment of causal pathways in nonlinear models," in *Causality: Statistical Perspectives and Applications*.

Pepe, M. and M. Mori (1993): "Kaplan-meier, marginal or conditional probability curves in summarizing competing risks failure time data?" *Statistics in Medicine*, 12, 737–751.

Pintilie, M. (2006): *Competing risks: A practical perspective*, Wiley.

Prentice, R. L. (2007): "Observational studies, clinical trials, and the women's health initiative," *Lifetime Data Analysis*, 13, 449–462.

Prentice, R. L., J. D. Kalbfleisch, A. V. Peterson, N. Flournoy, V. T. Farewell, and N. E. Breslow (1978): "Analysis of failure times in presence of competing risks," *Biometrics*, 34, 541–554.

Putter, H., M. Fiocco, and R. B. Geskus (2007): "Tutorial in biostatistics: competing risks and multi-state models," *Statistics in Medicine*, 26, 2389–2430, URL `http://dx.doi.org/10.1002/sim.2712`.

Rello, J., M. Rue, P. Jubert, G. Muses, R. Sonora, J. Valles, and M. S. Niederman (1997): "Survival in patients with nosocomial pneumonia: Impact of the severity of illness and the etiologic agent," *Critical Care Medicine*, 25, 1862–1867.

Resche-Rigon, M., E. Azoulay, and S. Chevret (2006a): "Evaluating mortality in intensive care units: contribution of competing risks analyses," *Critical Care*, 10, R5.

Resche-Rigon, M., E. Azoulay, and S. Chevret (2006b): "Evaluating mortality in intensive care units: contribution of competing risks analyses," *Critical Care*, 10, –.

Robins, J. (1986): "A new approach to causal inference in mortality studies with sustained exposure periods–application to control of the healthy worker survivor effect," *Mathematical Modelling*, 7, 1393–1512.

Robins, J., M. Hernán, and B. Brumback (2000): "Marginal structural models and causal inference in epidemiology," *Epidemiology*, 11, 550–560.

Robins, J. and Y. Ritov (1997): "Toward a curse of dimensionality appropriate (coda) asymptotic theory for semi-parametric models," *Statistics in Medicine*, 16, 285–319.

Robins, J. and A. Rotnitzky (1992): "Recovery of information and adjustment for dependent censoring using surrogate markers," *AIDS Epidemiology - Methodological Issues*, 29, 297–331.

Robins, J. and A. Rotnitzky (2001): "Inference for semiparametric models: Some questions and an answer - Comments," *Statistica Sinica*, 11, 920–936.

Robins, J., A. Rotnitzky, and S. D.0. (1999): "Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models," in M. Halloran and D. Berry, eds., *Statistical Models in Epidemiology: The Environment and Clinical Trials*, New-York: Springer-Verlag, 1–92.

Robins, J., M. Sued, Q. Lei-Gomez, and A. Rotnitzky (2008): "Performance of double-robust estimators when 'inverse probability' weights are highly variable," *Statistical Science*, 22, 544–559.

Robins, J. M. (1997): "Causal inference from complex longitudinal data," in *Latent variable modeling and applications to causality (Los Angeles, CA, 1994)*, *Lecture Notes in Statist.*, volume 120, New York: Springer, 69–117.

Robins, J. M. (1999): "Testing and estimation of direct effects by reparameterizing directed acyclic graphs with structural nested models," in *Computation, causation, and discovery*, Menlo Park, CA: AAAI Press, 349–405.

Robins, J. M. (2001): "Data, design, and background knowledge in etiologic inference," *Epidemiology*, 12, 313–320.

Robins, J. M. and S. Greenland (1986): "The role of model selection in causal inference from nonexperimental data," *American Journal of Epidemiology*, 123, 392–402.

Robins, J. M. and S. Greenland (1992): "Identifiability and exchangeability for direct and indirect effects." *Epidemiology*, 3, 143–155.

Robins, J. M., S. D. Mark, and N. W. K. (1992): "Estimating exposure effects by modelling the expectation of exposure conditional on confounders," *Biometrics*, 48, 479–495.

Robins, J. M. and A. Rotnitzky (2001): "Comment on a paper by p. bickel and j. kwon," *Statistica Sinica*, 11, 920–936.

Rosenbaum, P. R. (1984): "The consequences of adjustment for a concomitant variable that has been affected by the treatment," *Journal of the Royal Statistical Society Series a-Statistics in Society*, 147, 656–666.

Rosenbaum, P. R. and D. B. Rubin (1983): "The central role of the propensity score in observational studies for causal effects," *Biometrika*, 70, 41–55.

Rosenblum, M. and M. J. van der Laan (2009): "Using regression models to analyze randomized trials: Asymptotically valid hypothesis tests despite incorrectly specified models," *Biometrics*, 65, 937–945.

Rothman, M. and S. Greenland (1998): *Modern epidemiology 2nd ed.*., Philadelphia : Lippincott-Raven.

Rubin, D. (1978): "Bayesian inference for causal effects: the role of r and omization," *Annals Statistics*, 6, 34–58.

Rubin, D. B. (1987): *Multiple imputation for nonresponse in surveys*, Wiley series in probability and mathematical statistics Applied probability and statistics,, New York ;: Wiley.

Rubin, D. B. (1997): "Estimating causal effects from large data sets using propensity scores," *Annals of Internal Medicine*, 127, 757–763.

Safdar, N., C. Dezfulian, H. Collard, and S. Saint (2005): "Clinical and economic consequences of ventilator-associated pneumonia: a systematic review," *Critical Care Medicine*, 33, 2184–2193.

Satagopan, J. M., L. Ben-Porat, M. Berwick, M. Robson, D. Kutler, and A. D. Auerbach (2004): "A note on competing risks in survival data analysis," *British Journal of Cancer*, 91, 1229–1235.

Scharfstein, D., A. Rotnitzky, and J. Robins (1999): "Adjusting for non-ignorable drop-out using semiparametric non-response models." *Journal of the American Statistical Association*, 94, 1121–1146, with discussion and a rejoinder by the authors.

Schisterman, E. F., S. R. Cole, and R. W. Platt (2009): "Overadjustment bias and unnecessary adjustment in epidemiologic studies," *Epidemiology*, 20, 488–495.

Schulgen, G. and M. Schumacher (1996): "Estimation of prolongation of hospital stay attributable to nosocomial infections: new approaches based on multistate models," *Lifetime data anal*, 2, 219–240.

Schumacher, M., M. Wangler, M. Wolkewitz, and J. Beyersmann (2007): "Attributable mortality due to nosocomial infections - a simple and useful application of multistate models," *Methods of Information in Medicine*, 46, 595–600.

Schumacher, M., M. Wangler, M. Wolkewitz, and J. Beyersmann (2007): "Attributable mortality due to nosocomial infections: A simple and useful application of multistate models," *Methods of Information in Medicine* , 46, 595–600.

Shenassa, E. D., C. Daskalakis, A. Liebhaber, M. Braubach, and M. Brown (2007): "Dampness and mold in the home and depression: An examination of mold-related illness and perceived control of one's home as possible depression pathways," *American Journal of Public Health*, 97, 1893–1899.

Sjolander, A. (2009): "Propensity scores and M-structures," *Statistics in Medicine*, 28, 1416–1420.

Suetens, C., B. Jans, and H. Carsauw (1999): "Nosocomial infections in intensive care: results from the belgian national surveillance - 1996-1998," *Archives of Public Health*, 57, 221–231.

Suetens, C., A. Savey, A. Lepape, I. Morales, J. Carlet, and J. Fabry (2003): "Surveillance

of nosocomial infections in intensive care units: towards a consensual approach in europe," *Réanimation*, 12, 205–213.

Tan, Z. (2007): "Understanding OR, PS, and DR , Comment on "Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data" by Kang and Schafer," *Statistical Science*, 22, 560–568.

Tan, Z. (2008): "Understanding OR, PS, and DR," *Statistical Science*, 22, 560–568.

Tchetgen Tchetgen, E. J. and I. Shpitser (2011): "Semiparametric estimation of models for natural direct and indirect effects," Technical report, Harvard University.

Tian, J. and J. Pearl (2000): "Probabilities of causation: Bounds and identification." *Ann Math Artif Intel*, 28, 287–313.

Tibshirani, R. (1996): "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society, Series B*, 58, 267–288.

Timsit, J., S. Chevret, and J. Valcke (1996a): "Mortality of nosocomial pneumonia in ventilated patients: influence of diagnostic tools," *American Journal Respiratory Critocam Care Medicine*, 154, 116–123.

Timsit, J. F., S. Chevret, J. Valcke, B. Misset, B. Renaud, F. W. Goldstein, P. Vaury, and J. Carlet (1996b): "Mortality of nosocomial pneumonia in ventilated patients: Influence of diagnostic tools," *American Journal of Respiratory and Critical Care Medicine*, 154, 116–123.

Timsit, J. F., J. R. Zahar, and S. Chevret (2011): "Attributable mortality of ventilator-associated pneumonia," *Current Opinion in Critical Care*, 17, 464–71.

Tsiatis, A. (2006): *Semiparametric Theory and Missing Data*, New York: Springer.

Valls, J., A. Pobo, O. Garca-Esquirol, D. Mariscal, J. Real, and R. Fernndez (2007): "Excess ICU mortality attributable to ventilator-associated pneumonia: the role of early vs late onset." *Intensive Care Medicine*, 33, 1363–1368.

van der Laan, M. and J. Robins (2003): *Unified methods for censored longitudinal data and causality*, Springer and New York.

van der Laan, M. J. and M. L. Petersen (2008): "Direct effect models," *The International Journal of Biostatistics*, 4, Article 23.

Van Walraven, C., D. Davis, A. Forster, and G. Wells (2004): "Time-dependent bias was common in survival analyses published in leading clinical journals," *J of Clin Epidemiol.*, 57, 672–682.

VanderWeele, T. and S. Vansteelandt (2009): "Conceptual issues concerning mediation, interventions and composition. Statistics and its Interface," *in press*.

VanderWeele, T. J. (2009): "Concerning the consistency assumption in causal inference," *Epidemiology*, 20, 880–883 10.1097/EDE.0b013e3181bd5638.

VanderWeele, T. J. (2009): "Marginal structural models for the estimation of direct and indirect effects," *Epidemiology*, 20, 18–26.

VanderWeele, T. J. (2011): "Causal Mediation Analysis With Survival Data," *Epidemiology*, 22, 582–585.

VanderWeele, T. J. and S. Vansteelandt (2009): "Conceptual issues concerning mediation, interventions and composition," *Statistics and its Interface*, 2, 457–468.

VanderWeele, T. J. and S. Vansteelandt (2010): "Odds Ratios for Mediation Analysis for a Dichotomous Outcome," *American Journal of Epidemiology*, 172, 1339–1348.

Vansteelandt, S. (2009a): "Discussion on 'Identifiability and estimation of causal effects in randomized trials with noncompliance and completely non-ignorable missing-data'," *Biometrics*, 65, 686–689.

Vansteelandt, S. (2009b): "Estimating direct effects in cohort and case-control studies," *Epidemiology*, 20, 851–860.

Vansteelandt, S., M. Bekaert, and G. Claeskens (2010a): "On model selection and model misspecification in causal inference," *Statistical Methods in Medical Research*, journal article Statistical methods in medical research Stat Methods Med Res. 2010 Nov 12.

Vansteelandt, S., J. Carpenter, and M. Kenward (2010b): "Analysis of incomplete data using inverse probability weighting and doubly robust estimators," *Methodology*, 6, 37–48.

Vansteelandt, S., J. Carpenter, and M. G. Kenward (2010): "Analysis of Incomplete Data Using Inverse Probability Weighting and Doubly Robust Estimators," *Methodology*, 6, 37–48.

Vansteelandt, S., K. Mertens, C. Suetens, and E. Goetghebeur (2009): "Marginal structural models for partial exposure regimes," *Biostatistics*, 10, 46–59.

Vansteelandt, S., T. VanderWeele, E. J. Tchetgen, and J. M. Robins (2008): "Multiply robust inference for statistical interactions," *Journal of the American Statistical Association*, 103, 1693–1704.

Vincent, J. L., R. Moreno, J. Takala, S. Willatts, A. DeMendonca, H. Bruining, C. K. Reinhart, P. M. Suter, and L. G. Thijs (1996): "The sofa (sepsis-related organ failure assessment) score to describe organ dysfunction/failure," *Intensive Care Medicine*, 22, 707–710.

Vrijens, F., F. Hulstaert, B. Gordts, C. De Laet, S. Devriese, S. Van de Sande, M. Huybrechts, and G. Peeters (2009): "Infecties in België, deel II: Impact op Mortaliteit en Kosten. Health Services Research (HSR)." Technical report, Brussel: Federaal Kenniscentrum voor de Gezondheidszorg (KCE).

Wolkewitz, M., J. Beyersmann, P. Gastmeier, and M. Schumacher (2009): "Modeling the effect of time-dependent exposure on intensive care unit mortality," *Intensive Care Medicine*, 35, 826–832.

Wolkewitz, M., J. Beyersmann, and M. Schumacher (2010): "A note on statistical association

and causality derived from epidemiological icu data," *Intensive Care Medicine*, 36, 549–549, 554KV Times Cited:1 Cited References Count:4.

Wolkewitz, M., R. Vonberg, H. Grundmann, J. Beyersmann, P. Gastmeier, S. Barwolff, C. Geffers, M. Behnke, H. Ruden, and M. Schumacher (2008): "Risk factors for the development of nosocomial pneumonia and mortality on intensive care units: application of competing risks models," *Critical Care*, R44.

Wooldridge, J. (2009): "Should instrumental variables be used as matching variables?" Tech. rep., Michigan State University.

Wright, S. (1934): "The method of path coefficients," *Annals of Mathematical Statistics*, 5, 161–215.

Zou, H. and T. Hastie (2005): "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society, Series B*, 67, 301–320.